REPRESENTATION OF SINGULAR INTEGRALS BY DYADIC OPERATORS, AND THE A_2 THEOREM

TUOMAS P. HYTÖNEN

Department of Mathematics and Statistics
P.O.B. 68 (Gustaf Hällströmin katu 2b)
FI-00014 University of Helsinki, Finland
tuomas.hytonen@helsinki.fi

Abstract. This exposition presents a self-contained proof of the A_2 theorem, the quantitatively sharp norm inequality for singular integral operators in the weighted space $L^2(w)$. The strategy of the proof is a streamlined version of the author's original one, based on a probabilistic Dyadic Representation Theorem for singular integral operators. While more recent non-probabilistic approaches are also available now, the probabilistic method provides additional structural information, which has independent interest and other applications. The presentation emphasizes connections to the David–Journé $T(1)$ theorem, whose proof is obtained as a byproduct. Only very basic Probability is used; in particular, the conditional probabilities of the original proof are completely avoided.

Keywords: Singular integral, Calderón–Zygmund operator, weighted norm inequality, sharp estimate, A_2 theorem, $T(1)$ theorem

2010 *Mathematics Subject Classification.* 42B25, 42B35.

1. Introduction

The goal of this exposition is to prove the following A_2 *theorem*:

1.1. **Theorem.** *Let T be any Calderón–Zygmund operator on $\mathbb{R}^d$ (like the Hilbert transform on $\mathbb{R}$, the Beurling transform on $\mathbb{C} \simeq \mathbb{R}^2$, or any of the Riesz transforms in $\mathbb{R}^d$ for $d \geq 2$; see Section 3 for the general definition). Let $w : \mathbb{R}^d \to [0, \infty]$ be a weight in the Muckenhoupt class A_2, i.e.,*

$$[w]_{A_2} := \sup_Q \fint_Q w \cdot \fint_Q \frac{1}{w} < \infty \qquad \Big(\fint_Q w := \frac{1}{|Q|}\int_Q w\Big),$$

where the supremum is over all axes-parallel cubes Q in $\mathbb{R}^d$. Let $L^2(w)$ be the space of all measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that

$$\|f\|_{L^2(w)} := \Big(\int_{\mathbb{R}^d} |f|^2 w\Big)^{1/2} < \infty.$$

Then the following norm inequality is valid for any $f \in L^2(w)$, where C_T only depends on T and not on f or w:

$$\|Tf\|_{L^2(w)} \leq C_T \cdot [w]_{A_2} \cdot \|f\|_{L^2(w)}.$$

This general theorem for all Calderón–Zygmund operators is due to the author [15], but it was first obtained in the listed special cases by S. Petermichl and A. Volberg [40] and Petermichl [38, 39], and in various further particular instances by a number of others [3, 7, 24, 44]. See also Section 7.A for more details on the history of the problem.

Although several different proofs of Theorem 1.1 are known by now, I will present one that is a direct descendant of the original approach, but greatly streamlined in various places, based on ingredients from various subsequent proofs. On the large scale, I follow the strategy of my paper with C. Pérez, S. Treil and A. Volberg [21], the first simplification of my original proof [15]. This consists of the following steps, which have independent interest:

(1) Reduction to *dyadic shift operators* (the Dyadic Representation Theorem): every Calderón–Zygmund operator T has a (probabilistic) representation in terms of these simpler operators, and hence it suffices to prove a similar claim for every dyadic shift S in place of T. This was a key novelty of [15] when it first appeared. In this exposition, the probabilistic ingredients of this representation have been simplified from [15, 21], in that no conditional probabilities are needed.

(2) Reduction to *testing conditions* (a local $T(1)$ theorem): in order to have the full norm inequality

$$\|Sf\|_{L^2(w)} \leq C_S[w]_{A_2}\|f\|_{L^2(w)},$$

it suffices to have such an inequality for special test functions only:

$$\|S(1_Q w^{-1})\|_{L^2(w)} \leq C_S[w]_{A_2}\|1_Q w^{-1}\|_{L^2(w)},$$
$$\|S^*(1_Q w)\|_{L^2(w^{-1})} \leq C_S[w]_{A_2}\|1_Q w\|_{L^2(w^{-1})}.$$

This goes essentially back to F. Nazarov, Treil and Volberg [34]. (In the original proof [15], in contrast to the simplification [21], this reduction was done on the level of the Calderón–Zygmund operator, using a more difficult variant due to Pérez, Treil and Volberg [37]).

(3) Verification of the testing conditions for S. This was first achieved by M. T. Lacey, Petermichl and M. C. Reguera [24], although some adjustments were necessary to achieve the full generality in [15].

As said, several different proofs and extensions of the A_2 theorem have appeared over the past few years; see the final section for further discussion and references. In particular, it is now known that the probabilistic Dyadic Representation Theorem may be replaced by a deterministic Dyadic Domination Theorem. Its first version, a domination in norm, is due to A. Lerner [26], and based on his clever local oscillation formula [25]; this was subsequently improved to pointwise domination by J. M. Conde-Alonso and G. Rey [2] and, independently, by Lerner and Nazarov [29]. Yet another approach to the pointwise domination was found by Lacey [23] and again simplified by Lerner [28]; this has the virtue of covering the biggest class of operators admissible for the A_2 theorem at the present state of knowledge. However, the probabilistic method continues to have its independent interest: it achieves the reduction to dyadic model operators as a linear *identity*, in contrast to the (non-linear) *upper bound* provided the deterministic domination. As such, it provides a structure theorem for singular integral operators, which has found other uses beyond the weighted norm inequalities, including the following:

- The *theorem itself* is applied to the estimation of *commutators* of Calderón–Zygmund operators and BMO functions in a multi-parameter setting by L. Dalenc and Y. Ou [4] and in a two-weight setting by I. Holmes, M. Lacey and B. Wick [13, 14]; it is also applied to sharp norm bounds for *vector-valued extensions* of Calderón–Zygmund operators by S. Pott and A. Stoica [41].
- The *methods behind this theorem* have been generalized by H. Martikainen [30] and Y. Ou [36] to the analysis of *bi-parameter singular integrals*, yielding new $T(1)$ and $T(b)$ type theorems for these operators.

Whereas the domination method *assumes* the unweighted L^2 boundedness of the operator T, the representation method can (and will, in this exposition) be set up in such a way that it *derives* the unweighted boundedness from a priori weaker assumptions as a byproduct. Indeed, a proof of the $T(1)$ theorem of G. David and J.-L. Journé [5] is obtained as a byproduct of the present exposition, and this approach was lifted to the nontrivial case of bi-parameter singular integrals in the mentioned works of Martikainen [30] and Ou [36]. Of course, the deterministic domination method has its own advantages, but the point that I want to make here is that so does the probabilistic approach, which I present in the following exposition.

2. Preliminaries

The standard (or reference) system of dyadic cubes is

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

We will need several dyadic systems, obtained by translating the reference system as follows. Let $\omega = (\omega_j)_{j\in\mathbb{Z}} \in (\{0,1\}^d)^{\mathbb{Z}}$ and

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

Then

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

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

2.A. **Haar functions.** Any given dyadic system $\mathscr{D}$ has a natural function system associated to it: the Haar functions. In one dimension, there are two Haar functions associated with an interval I: the non-cancellative $h_I^0 := |I|^{-1/2} 1_I$ and the cancellative $h_I^1 := |I|^{-1/2}(1_{I_\ell} - 1_{I_r})$, where I_ℓ and I_r are the left and right halves of I. In d dimensions, the Haar functions on a cube $I = I_1 \times \cdots \times I_d$ are formed of all the products of the one-dimensional Haar functions:

$$h_I^\eta(x) = h_{I_1 \times \cdots \times I_d}^{(\eta_1, \ldots, \eta_d)}(x_1, \ldots, x_d) := \prod_{i=1}^d h_{I_i}^{\eta_i}(x_i).$$

The non-cancellative $h_I^0 = |I|^{-1/2} 1_I$ has the same formula as in $d = 1$. All other $2^d - 1$ Haar functions h_I^η with $\eta \in \{0,1\}^d \setminus \{0\}$ are cancellative, i.e., satisfy $\int h_I^\eta = 0$, since they are cancellative in at least one coordinate direction.

For a fixed $\mathscr{D}$, all the cancellative Haar functions h_I^η, $I \in \mathscr{D}$ and $\eta \in \{0,1\}^d \setminus \{0\}$, form an orthonormal basis of $L^2(\mathbb{R}^d)$. Hence any function $f \in L^2(\mathbb{R}^d)$ has the orthogonal expansion

$$f = \sum_{I \in \mathscr{D}} \sum_{\eta \in \{0,1\}^d \setminus \{0\}} \langle f, h_I^\eta \rangle h_I^\eta.$$

Since the different η's seldom play any major role, this will be often abbreviated (with slight abuse of language) simply as

$$f = \sum_{I \in \mathscr{D}} \langle f, h_I \rangle h_I,$$

and the summation over η is understood implicitly.

2.B. **Dyadic shifts.** A dyadic shift with parameters $i, j \in \mathbb{N} := \{0, 1, 2, \ldots\}$ is an operator of the form

$$Sf = \sum_{K \in \mathscr{D}} A_K f, \qquad A_K f = \sum_{\substack{I, J \in \mathscr{D}; I, J \subseteq K \\ \ell(I) = 2^{-i} \ell(K) \\ \ell(J) = 2^{-j} \ell(K)}} a_{IJK} \langle f, h_I \rangle h_J,$$

where h_I is a Haar function on I (similarly h_J), and the a_{IJK} are coefficients with

$$|a_{IJK}| \leq \frac{\sqrt{|I||J|}}{|K|}.$$

It is also required that all subshifts

$$S_{\mathscr{Q}} = \sum_{K \in \mathscr{Q}} A_K, \qquad \mathscr{Q} \subseteq \mathscr{D},$$

map $S_{\mathscr{Q}} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ with norm at most one.

The shift is called cancellative, if all the h_I and h_J are cancellative; otherwise, it is called non-cancellative.

The notation A_K indicates an "averaging operator" on K. Indeed, from the normalization of the Haar functions, it follows that

$$|A_K f| \leq 1_K \fint_K |f|$$

pointwise.

For cancellative shifts, the L^2 boundedness is automatic from the other conditions. This is a consequence of the following facts:

- The pointwise bound for each A_K implies that $\|A_K f\|_{L^p} \leq \|f\|_{L^p}$ for all $p \in [1, \infty]$; in particular, these components of S are uniformly bounded on L^2 with norm one. (This first point is true even in the non-cancellative case.)
- Let $\mathbb{D}^i_K$ denote the orthogonal projection of L^2 onto $\operatorname{span}\{h_I : I \subseteq K, \ell(I) = 2^{-i}\ell(K)\}$. When i is fixed, it follows readily that any two $\mathbb{D}^i_K$ are orthogonal to each other. (This depends on the use of cancellative h_I.) Moreover, we have $A_K = \mathbb{D}^j_K A_K \mathbb{D}^i_K$. Then the boundedness of S follows from two applications of Pythagoras' theorem with the uniform boundedness of the A_K in between.

A prime example of a non-cancellative shift (and the only one we need in these lectures) is the *dyadic paraproduct*

$$\Pi_b f = \sum_{K \in \mathscr{D}} \langle b, h_K \rangle \langle f \rangle_K h_K = \sum_{K \in \mathscr{D}} |K|^{-1/2} \langle b, h_K \rangle \cdot \langle f, h^0_K \rangle h_K,$$

where $b \in \mathrm{BMO}_d$ (the dyadic BMO space) and h_K is a cancellative Haar function. This is a dyadic shift with parameters $(i, j) = (0, 0)$, where $a_{IJK} = |K|^{-1/2}\langle b, h_K\rangle$ for $I = J = K$. The L^2 boundedness of the paraproduct, if and only if $b \in \mathrm{BMO}_d$, is part of the classical theory. Actually, to ensure the normalization condition of the shift, it should be further required that $\|b\|_{\mathrm{BMO}_d} \leq 1$.

2.C. **Random dyadic systems; good and bad cubes.** We obtain a notion of *random dyadic systems* by equipping the parameter set $\Omega := (\{0,1\}^d)^{\mathbb{Z}}$ with the natural probability measure: each component ω_j has an equal probability 2^{-d} of taking any of the 2^d values in $\{0,1\}^d$, and all components are independent of each other.

Let $\phi : [0,1] \to [0,1]$ be a fixed *modulus of continuity*: a strictly increasing function with $\phi(0) = 0$, $\phi(1) = 1$, and $t \mapsto \phi(t)/t$ decreasing (hence $\phi(t) \geq t$) with $\lim_{t\to 0} \phi(t)/t = \infty$. We further require the *Dini condition*

$$\int_0^1 \phi(t) \frac{\mathrm{d}t}{t} < \infty.$$

Main examples include $\phi(t) = t^\gamma$ with $\gamma \in (0,1)$ and

$$\phi(t) = \Big(1 + \frac{1}{\gamma} \log \frac{1}{t}\Big)^{-\gamma}, \qquad \gamma > 1.$$

We also fix a (large) parameter $r \in \mathbb{Z}_+$. (How large, will be specified shortly.)

A cube $I \in \mathscr{D}^\omega$ is called bad if there exists $J \in \mathscr{D}^\omega$ such that $\ell(J) \geq 2^r \ell(I)$ and

$$\operatorname{dist}(I, \partial J) \leq \phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\ell(J):$$

roughly, I is relatively close to the boundary of a much bigger cube.

2.1. *Remark.* This definition of good cubes goes back to Nazarov–Treil–Volberg [33] in the context of singular integrals with respect to non-doubling measures. They used the modulus of continuity $\phi(t) = t^\gamma$, where γ was chosen to depend on the dimension and the Hölder exponent of the Calderón–Zygmund kernel via

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

This choice has become "canonical" in the subsequent literature, including the original proof of the A_2 theorem. However, other choices can also be made, as we do here.

We make some basic probabilistic observations related to badness. Let $I \in \mathscr{D}^0$ be a reference interval. The *position* of the translated interval

$$I\dot{+}\omega = I + \sum_{j:2^{-j}<\ell(I)} 2^{-j}\omega_j,$$

by definition, depends only on ω_j for $2^{-j} < \ell(I)$. On the other hand, the *badness* of $I\dot{+}\omega$ depends on its *relative position* with respect to the bigger intervals

$$J\dot{+}\omega = J + \sum_{j:2^{-j}<\ell(I)} 2^{-j}\omega_j + \sum_{j:\ell(I)\leq 2^{-j}<\ell(I)} 2^{-j}\omega_j.$$

The same translation component $\sum_{j:2^{-j}<\ell(I)} 2^{-j}\omega_j$ appears in both $I\dot{+}\omega$ and $J\dot{+}\omega$, and so does not affect the relative position of these intervals. Thus this relative position, and hence the badness of I, depends only on ω_j for $2^{-j} \geq \ell(I)$. In particular:

2.2. **Lemma.** *For $I \in \mathscr{D}^0$, the position and badness of $I\dot{+}\omega$ are independent random variables.*

Another observation is the following: by symmetry and the fact that the condition of badness only involves relative position and size of different cubes, it readily follows that the probability of a particular cube $I\dot{+}\omega$ being bad is equal for all cubes $I \in \mathscr{D}^0$:

$$\mathbb{P}_\omega(I\dot{+}\omega \text{ bad}) = \pi_{\text{bad}} = \pi_{\text{bad}}(r, d, \phi).$$

The final observation concerns the value of this probability:

2.3. **Lemma.** *We have*

$$\pi_{\text{bad}} \leq 8d \int_0^{2^{-r}} \phi(t)\frac{\mathrm{d}t}{t};$$

in particular, $\pi_{\text{bad}} < 1$ if $r = r(d,\phi)$ is chosen large enough.

With $r = r(d, \phi)$ chosen like this, we then have $\pi_{\text{good}} := 1 - \pi_{\text{bad}} > 0$, namely, good situations have positive probability!

Proof. Observe that in the definition of badness, we only need to consider those J with $I \subseteq J$. Namely, if I is close to the boundary of some bigger J, we can always find another dyadic J' of the same size as J which contains I, and then I will also be close to the boundary of J'. Hence we need to consider the relative position of I with respect to each $J \supset I$ with $\ell(J) = 2^k\ell(I)$ and $k = r, r+1, \ldots$ For a fixed k, this relative position is determined by

$$\sum_{j:\ell(I)\leq 2^{-j}<2^k\ell(I)} 2^{-j}\omega_j,$$

which has 2^{kd} different values with equal probability. These correspond to the subcubes of J of size $\ell(I)$.

Now bad position of I are those which are within distance $\phi(\ell(I)/\ell(J)) \cdot \ell(J)$ from the boundary. Since the possible position of the subcubes are discrete, being integer multiples of $\ell(I)$, the effective bad boundary region has depth

$$\begin{aligned}\Big\lceil \phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\frac{\ell(J)}{\ell(I)}\Big\rceil \ell(I) &\leq \Big(\phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\frac{\ell(J)}{\ell(I)} + 1\Big)\ell(I) \\ &= \ell(J)\Big(\phi\Big(\frac{\ell(I)}{\ell(J)}\Big) + \frac{\ell(I)}{\ell(J)}\Big) \leq 2\ell(J)\phi\Big(\frac{\ell(I)}{\ell(J)}\Big),\end{aligned}$$

by using that $t \leq \phi(t)$.

The good region is the cube inside J, whose side-length is $\ell(J)$ minus twice the depth of the bad boundary region:

$$\ell(J) - 2\Big\lceil \phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\frac{\ell(J)}{\ell(I)}\Big\rceil \ell(I) \geq \ell(J) - 4\ell(J)\phi\Big(\frac{\ell(I)}{\ell(J)}\Big).$$

Hence the volume of the bad region is

$$\begin{aligned}|J| - \Big(\ell(J) - 2\Big\lceil \phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\frac{\ell(J)}{\ell(I)}\Big\rceil \ell(I)\Big)^d &\leq |J|\Big(1 - \Big(1 - 4\phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\Big)^d\Big) \\ &\leq |J| \cdot 4d\phi\Big(\frac{\ell(I)}{\ell(J)}\Big)\end{aligned}$$

by the elementary inequality $(1-\alpha)^d \geq 1 - \alpha d$ for $\alpha \in [0,1]$. (We assume that r is at least so large that $4\phi(2^{-r}) \leq 1$.)

So the fraction of the bad region of the total volume is at most $4d\phi(\ell(I)/\ell(J)) = 4d\phi(2^{-k})$ for a fixed $k = r, r+1, \ldots$. This gives the final estimate

$$\begin{aligned}\mathbb{P}_\omega(I \dot{+} \omega \text{ bad}) &\leq \sum_{k=r}^{\infty} 4d\phi(2^{-k}) = \sum_{k=r}^{\infty} 8d\frac{\phi(2^{-k})}{2^{-k}}2^{-k-1} \\ &\leq \sum_{k=r}^{\infty} 8d\int_{2^{-k-1}}^{2^{-k}} \frac{\phi(t)}{t}\,\mathrm{d}t = 8d\int_0^{2^{-r}} \phi(t)\frac{\mathrm{d}t}{t},\end{aligned}$$

where we used that $\phi(t)/t$ is decreasing in the last inequality. □

3. The dyadic representation theorem

Let T be a Calderón–Zygmund operator on $\mathbb{R}^d$. That is, it acts on a suitable dense subspace of functions in $L^2(\mathbb{R}^d)$ (for the present purposes, this class should at least contain the indicators of cubes in $\mathbb{R}^d$) and has the kernel representation

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

Moreover, the kernel should satisfy the *standard estimates*, which we here assume in a slightly more general form than usual, involving another modulus of continuity ψ, like the one considered above:

$$\begin{aligned}|K(x,y)| &\leq \frac{C_0}{|x-y|^d}, \\ |K(x,y) - K(x',y)| + |K(y,x) - K(y,x')| &\leq \frac{C_\psi}{|x-y|^d}\psi\Big(\frac{|x-x'|}{|x-y|}\Big)\end{aligned}$$

for all $x, x', y \in \mathbb{R}^d$ with $|x-y| > 2|x-x'|$. Let us denote the smallest admissible constants C_0 and C_ψ by $\|K\|_{CZ_0}$ and $\|K\|_{CZ_\psi}$. The classical standard estimates correspond to the choice $\psi(t) = t^\alpha$, $\alpha \in (0,1]$, in which case we write $\|K\|_{CZ_\alpha}$ for $\|K\|_{CZ_\psi}$.

We say that T is a bounded Calderón–Zygmund operator, if in addition $T : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, and we denote its operator norm by $\|T\|_{L^2 \to L^2}$.

Let us agree that $|\ |$ stands for the ℓ^∞ norm on $\mathbb{R}^d$, i.e., $|x| := \max_{1 \leq i \leq d} |x_i|$. While the choice of the norm is not particularly important, this choice is slightly more convenient than the usual Euclidean norm when dealing with cubes as we will: e.g., the diameter of a cube in the ℓ^∞ norm is equal to its sidelength $\ell(Q)$.

Let us first formulate the dyadic representation theorem for general moduli of continuity, and then specialize it to the usual standard estimates. Define the following coefficients for $i, j \in \mathbb{N}$:

$$\tau(i,j) := \phi(2^{-\max\{i,j\}})^{-d} \psi\big(2^{-\max\{i,j\}} \phi(2^{-\max\{i,j\}})^{-1}\big),$$

if $\min\{i,j\} > 0$; and

$$\tau(i,j) := \Psi\big(2^{-\max\{i,j\}} \phi(2^{-\max\{i,j\}})^{-1}\big), \qquad \Psi(t) := \int_0^t \psi(s) \frac{\mathrm{d}s}{s},$$

if $\min\{i,j\} = 0$.

We assume that ϕ and ψ are such, that

$$\sum_{i,j=0}^{\infty} \tau(i,j) \eqsim \int_0^1 \frac{1}{\phi(t)^d} \psi\Big(\frac{t}{\phi(t)}\Big) \frac{\mathrm{d}t}{t} + \int_0^1 \Psi\Big(\frac{t}{\phi(t)}\Big) \frac{\mathrm{d}t}{t} < \infty. \tag{3.1}$$

This is the case, in particular, when $\psi(t) = t^\alpha$ (usual standard estimates) and $\phi(t) = (1 + a^{-1} \log t^{-1})^{-\gamma}$; then one checks that

$$\tau(i,j) \lesssim P(\max\{i,j\}) 2^{-\alpha \max\{i,j\}}, \qquad P(j) = (1+j)^{\gamma(d+\alpha)},$$

which clearly satisfies the required convergence. However, it is also possible to treat weaker forms of the standard estimates with a logarithmic modulus $\psi(t) = (1 + a^{-1} \log t^{-1})^{-\alpha}$. This might be of some interest for applications, but we do not pursue this line any further here.

3.2. **Theorem.** *Let T be a bounded Calderón–Zygmund operator with modulus of continuity satisfying the above assumption. Then it has an expansion, say for $f, g \in C_c^1(\mathbb{R}^d)$,*

$$\langle g, Tf \rangle = c \cdot \big(\|T\|_{L^2 \to L^2} + \|K\|_{CZ_\psi}\big) \cdot \mathbb{E}_\omega \sum_{i,j=0}^{\infty} \tau(i,j) \langle g, S_\omega^{ij} f \rangle,$$

where c is a dimensional constant and S_ω^{ij} is a dyadic shift of parameters (i,j) on the dyadic system $\mathscr{D}^\omega$; all of them except possibly S_ω^{00} are cancellative.

The first version of this theorem appeared in [15], and another one in [21]. The present proof is yet another variant of the same argument. It is slightly simpler in terms of the probabilistic tools that are used: no conditional probabilities are needed, although they were important for the original arguments.

In proving this theorem, we do not actually need to employ the full strength of the assumption that $T : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$; rather it suffices to have the kernel conditions plus the following conditions of the $T1$ theorem of David–Journé:

$$|\langle 1_Q, T1_Q\rangle| \le C_{WBP}|Q| \quad \text{(weak boundedness property)},$$
$$T1 \in \mathrm{BMO}(\mathbb{R}^d), \quad T^*1 \in \mathrm{BMO}(\mathbb{R}^d).$$

Let us denote the smallest C_{WBP} by $\|T\|_{WBP}$. Then we have the following more precise version of the representation:

3.3. **Theorem.** *Let T be a Calderón–Zygmund operator with modulus of continuity satisfying the above assumption. Then it has an expansion, say for $f, g \in C_c^1(\mathbb{R}^d)$,*

$$\begin{aligned}\langle g, Tf\rangle &= c\cdot\big(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}\big)\mathbb{E}_\omega \sum_{\substack{i,j=0\\ \max\{i,j\}>0}}^{\infty} \tau(i,j)\langle g, S^{ij}_\omega f\rangle \\ &\quad + c\cdot\big(\|K\|_{CZ_0} + \|T\|_{WBP}\big)\mathbb{E}_\omega\langle g, S^{00}_\omega f\rangle + \mathbb{E}_\omega\langle g, \Pi^\omega_{T1} f\rangle + \mathbb{E}_\omega\langle g, (\Pi^\omega_{T^*1})^* f\rangle\end{aligned}$$

where S^{ij}_ω is a cancellative dyadic shift of parameters (i,j) on the dyadic system $\mathscr{D}^\omega$, and Π^ω_b is a dyadic paraproduct on the dyadic system $\mathscr{D}^\omega$ associated with the BMO*-function $b \in \{T1, T^*1\}$.*

3.4. *Remark.* Note that $\Pi^\omega_b = \|b\|_{\mathrm{BMO}}\cdot S^\omega_b$, where $S^\omega_b = \Pi^\omega_b/\|b\|_{\mathrm{BMO}}$ is a shift with the correct normalization. Hence, writing everything in terms of normalized shifts, as in Theorem 3.2, we get the factor $\|T1\|_{\mathrm{BMO}} \lesssim \|T\|_{L^2\to L^2} + \|K\|_{CZ_\psi}$ in the second-to-last term, and $\|T^*1\|_{\mathrm{BMO}} \lesssim \|T\|_{L^2\to L^2} + \|K\|_{CZ_\psi}$ in the last one. The proof will also show that both occurrences of the factor $\|K\|_{CZ_0}$ could be replaced by $\|T\|_{L^2\to L^2}$, giving the statement of Theorem 3.2 (since trivially $\|T\|_{WBP} \le \|T\|_{L^2\to L^2}$).

As a by-product, Theorem 3.3 delivers a proof of the $T1$ theorem: under the above assumptions, the operator T is already bounded on $L^2(\mathbb{R}^d)$. Namely, all the dyadic shifts S^{ij}_ω are uniformly bounded on $L^2(\mathbb{R}^d)$ by definition, and the convergence condition (3.1) ensures that so is their average representing the operator T. This by-product proof of the $T1$ theorem is not a coincidence, since the proof of Theorems 3.2 and 3.3 was actually inspired by the proof of the $T1$ theorem for non-doubling measures due to Nazarov–Treil–Volberg [33] and its vector-valued extension [16].

A key to the proof of the dyadic representation is a random expansion of T in terms of Haar functions h_I, where the bad cubes are avoided:

3.5. **Proposition.**

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

where

$$\mathrm{smaller}\{I,J\} := \begin{cases} I & \text{if } \ell(I)\le\ell(J),\\ J & \text{if } \ell(J)>\ell(I).\end{cases}$$

Proof. Recall that

$$f = \sum_{I\in\mathscr{D}^0}\langle f, h_{I\dot+\omega}\rangle h_{I\dot+\omega}$$

for any fixed $\omega\in\Omega$; and we can also take the expectation $\mathbb{E}_\omega$ of both sides of this identity.

Let

$$1_{\text{good}}(I\dot{+}\omega) := \begin{cases} 1, & \text{if } I\dot{+}\omega \text{ is good,} \\ 0, & \text{else} \end{cases}$$

We make use of the above random Haar expansion of f, multiply and divide by

$$\pi_{\text{good}} = \mathbb{P}_\omega(I\dot{+}\omega\,\text{good}) = \mathbb{E}_\omega 1_{\text{good}}(I\dot{+}\omega),$$

and use the independence from Lemma 2.2 to get:

$$\begin{aligned} \langle g, Tf\rangle &= \mathbb{E}_\omega \sum_I \langle g, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle \\ &= \frac{1}{\pi_{\text{good}}} \sum_I \mathbb{E}_\omega[1_{\text{good}}(I\dot{+}\omega)]\mathbb{E}_\omega[\langle g, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle] \\ &= \frac{1}{\pi_{\text{good}}}\mathbb{E}_\omega \sum_I 1_{\text{good}}(I\dot{+}\omega)\langle g, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle \\ &= \frac{1}{\pi_{\text{good}}}\mathbb{E}_\omega \sum_{I,J} 1_{\text{good}}(I\dot{+}\omega)\langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle. \end{aligned}$$

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

$$\begin{aligned} &\frac{1}{\pi_{\text{good}}} \sum_{\ell(I)>\ell(J)} \mathbb{E}_\omega[1_{\text{good}}(I\dot{+}\omega)\langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle] \\ &= \frac{1}{\pi_{\text{good}}} \sum_{\ell(I)>\ell(J)} \mathbb{E}_\omega[1_{\text{good}}(I\dot{+}\omega)]\mathbb{E}_\omega[\langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle] \\ &= \mathbb{E}_\omega \sum_{\ell(I)>\ell(J)} \langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle, \end{aligned}$$

and hence

$$\begin{aligned} \langle g, Tf\rangle = \frac{1}{\pi_{\text{good}}}\mathbb{E}_\omega &\sum_{\ell(I)\leq\ell(J)} 1_{\text{good}}(I\dot{+}\omega)\langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle \\ &+ \mathbb{E}_\omega \sum_{\ell(I)>\ell(J)} \langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle. \end{aligned}$$

Comparison with the basic identity

$$\langle g, Tf\rangle = \mathbb{E}_\omega \sum_{I,J} \langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle \tag{3.6}$$

shows that

$$\begin{aligned} &\mathbb{E}_\omega \sum_{\ell(I)\leq\ell(J)} \langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle \\ &\quad= \frac{1}{\pi_{\text{good}}}\mathbb{E}_\omega \sum_{\ell(I)\leq\ell(J)} 1_{\text{good}}(I\dot{+}\omega)\langle g, h_{J\dot{+}\omega}\rangle\langle h_{J\dot{+}\omega}, Th_{I\dot{+}\omega}\rangle\langle h_{I\dot{+}\omega}, f\rangle. \end{aligned}$$

Symmetrically, we also have

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

and this completes the proof. $\square$

This is essentially the end of probability in this proof. Henceforth, we can simply concentrate on the summation inside $\mathbb{E}_\omega$, for a fixed value of $\omega \in \Omega$, and manipulate it into the required form. Moreover, we will concentrate on the half of the sum with $\ell(J) \geq \ell(I)$, the other half being handled symmetrically. We further divide this sum into the following parts:

$$\begin{aligned}\sum_{\ell(I)\leq\ell(J)} &= \sum_{\mathrm{dist}(I,J)>\ell(J)\phi(\ell(I)/\ell(J))} + \sum_{I\subsetneq J} + \sum_{I=J} + \sum_{\substack{\mathrm{dist}(I,J)\leq\ell(J)\phi(\ell(I)/\ell(J))\\ I\cap J=\varnothing}} \\ &=: \sigma_{\mathrm{out}} + \sigma_{\mathrm{in}} + \sigma_{=} + \sigma_{\mathrm{near}}.\end{aligned}$$

In order to recognize these series as sums of dyadic shifts, we need to locate, for each pair (I, J) appearing here, a common dyadic ancestor which contains both of them. The existence of such containing cubes, with control on their size, is provided by the following:

3.7. **Lemma.** *If $I \in \mathscr{D}$ is good and $J \in \mathscr{D}$ is a disjoint ($J \cap I = \varnothing$) cube with $\ell(J) \geq \ell(I)$, then there exists $K \supseteq I \cup J$ which satisfies*

$$\begin{aligned}\ell(K) \leq 2^r \ell(I), \qquad &\textit{if} \qquad \mathrm{dist}(I,J) \leq \ell(J)\phi\Big(\frac{\ell(I)}{\ell(J)}\Big),\\ \ell(K)\phi\Big(\frac{\ell(I)}{\ell(K)}\Big) \leq 2^r \,\mathrm{dist}(I,J), \qquad &\textit{if} \qquad \mathrm{dist}(I,J) > \ell(J)\phi\Big(\frac{\ell(I)}{\ell(J)}\Big).\end{aligned}$$

Proof. Let us start with the following initial observation: if $K \in \mathscr{D}$ satisfies $I \subseteq K$, $J \subset K^c$, and $\ell(K) \geq 2^r \ell(I)$, then

$$\ell(K)\phi\Big(\frac{\ell(I)}{\ell(K)}\Big) < \mathrm{dist}(I, \partial K) = \mathrm{dist}(I, K^c) \leq \mathrm{dist}(I, J).$$

Case $\mathrm{dist}(I,J) \leq \ell(J)\phi(\ell(I)/\ell(J))$. As $I \cap J = \varnothing$, we have $\mathrm{dist}(I,J) = \mathrm{dist}(I, \partial J)$, and since I is good, this implies $\ell(J) < 2^r \ell(I)$. Let $K = I^{(r)}$, and assume for contradiction that $J \subset K^c$. Then the initial observation implies that

$$\ell(K)\phi\Big(\frac{\ell(I)}{\ell(K)}\Big) < \mathrm{dist}(I,J) \leq \ell(J)\phi\Big(\frac{\ell(I)}{\ell(J)}\Big).$$

Dividing both sides by $\ell(I)$ and recalling that $\phi(t)/t$ is decreasing, this implies that $\ell(K) < \ell(J)$, a contradiction with $\ell(K) = 2^r \ell(I) > \ell(J)$. Hence $J \not\subset K^c$, and since $\ell(J) < \ell(K)$, this implies that $J \subset K$.

Case $\mathrm{dist}(I,J) > \ell(J)\phi(\ell(I)/\ell(J))$. Consider the minimal $K \supset I$ with $\ell(K) \geq 2^r \ell(I)$ and $\mathrm{dist}(I,J) \leq \ell(K)\phi(\ell(I)/\ell(K))$. (Since $\phi(t)/t \to \infty$ as $t \to 0$, this bound holds for all large enough K.) Then (since $\phi(t)/t$ is decreasing) $\ell(K) > \ell(J)$, and by the initial observation, $J \not\subset K^c$. Hence either $J \subset K$, and it suffices to estimate $\ell(K)$.

By the minimality of K, there holds at least one of

$$\tfrac{1}{2}\ell(K) < 2^r \ell(I) \qquad \text{or} \qquad \tfrac{1}{2}\ell(K)\phi\Big(\frac{\ell(I)}{\frac{1}{2}\ell(K)}\Big) < \mathrm{dist}(I,J),$$

and the latter immediately implies that $\ell(K)\phi(\ell(I)/\ell(K)) < 2\operatorname{dist}(I,J)$. In the first case, since $\ell(I) \le \ell(J) \le \ell(K)$, we have

$$\ell(K)\phi\Big(\frac{\ell(I)}{\ell(K)}\Big) \le 2^r \ell(I)\Big(\frac{\ell(I)}{\ell(K)}\Big) \le 2^r \ell(J)\Big(\frac{\ell(I)}{\ell(J)}\Big) < 2^r \operatorname{dist}(I,J),$$

so the required bound is true in each case. □

We denote the minimal such K by $I \vee J$, thus

$$I \vee J := \bigcap_{K \supseteq I \cup J} K.$$

3.A. **Separated cubes, σ_{out}.** We reorganize the sum σ_{out} with respect to the new summation variable $K = I \vee J$, as well as the relative size of I and J with respect to K:

$$\sigma_{\mathrm{out}} = \sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\sum_{K} \sum_{\substack{\operatorname{dist}(I,J) > \ell(J)\phi(\ell(I)/\ell(J)) \\ I \vee J = K \\ \ell(I) = 2^{-i}\ell(K), \ell(J) = 2^{-j}\ell(K)}} .$$

Note that we can start the summation from 1 instead of 0, since the disjointness of I and J implies that $K = I \vee J$ must be strictly larger than either of I and J. The goal is to identify the quantity in parentheses as a decaying factor times a cancellative averaging operator with parameters (i,j).

3.8. **Lemma.** *For I and J appearing in σ_{out}, we have*

$$|\langle h_J, Th_I\rangle| \lesssim \|K\|_{CZ_\psi} \frac{\sqrt{|I||J|}}{|K|}\phi\Big(\frac{\ell(I)}{\ell(K)}\Big)^{-d}\psi\Big(\frac{\ell(I)}{\ell(K)}\phi\Big(\frac{\ell(I)}{\ell(K)}\Big)^{-1}\Big), \qquad K = I \vee J.$$

Proof. Using the cancellation of h_I, standard estimates, and Lemma 3.7

$$\begin{aligned}
|\langle h_J, Th_I\rangle| &= \Big| \iint h_J(x)K(x,y)h_I(y)\,\mathrm{d}y\,\mathrm{d}x \Big| \\
&= \Big| \iint h_J(x)[K(x,y) - K(x,y_I)]h_I(y)\,\mathrm{d}y\,\mathrm{d}x \Big| \\
&\lesssim \|K\|_{CZ_\psi} \iint |h_J(x)| \frac{1}{\operatorname{dist}(I,J)^d}\psi\Big(\frac{\ell(I)}{\operatorname{dist}(I,J)}\Big)|h_I(y)|\,\mathrm{d}y\,\mathrm{d}x \\
&= \|K\|_{CZ_\psi} \frac{1}{\operatorname{dist}(I,J)^d}\psi\Big(\frac{\ell(I)}{\operatorname{dist}(I,J)}\Big)\|h_J\|_1\|h_I\|_1 \\
&\lesssim \|K\|_{CZ_\psi} \frac{1}{\ell(K)^d}\phi\Big(\frac{\ell(I)}{\ell(K)}\Big)^{-d}\psi\Big(\frac{\ell(I)}{\ell(K)}\phi\Big(\frac{\ell(I)}{\ell(K)}\Big)^{-1}\Big)\sqrt{|J|}\sqrt{|I|}.
\end{aligned}$$
□

3.9. **Lemma.**

$$\begin{aligned}
\sum_{\substack{\operatorname{dist}(I,J) > \ell(J)\phi(\ell(I)/\ell(J)) \\ I \vee J = K \\ \ell(I) = 2^{-i}\ell(K) \le \ell(J) = 2^{-j}\ell(K)}} & 1_{\mathrm{good}}(I)\cdot\langle g, h_J\rangle\langle h_J, Th_I\rangle\langle h_I, f\rangle \\
&= \|K\|_{CZ_\psi}\phi(2^{-i})^{-d}\psi\big(2^{-i}\phi(2^{-i})^{-1}\big)\langle g, A_K^{ij} f\rangle,
\end{aligned}$$

where A_K^{ij} is a cancellative averaging operator with parameters (i,j).

Proof. By the previous lemma, substituting $\ell(I)/\ell(K) = 2^{-i}$,

$$|\langle h_J, Th_I\rangle| \lesssim \|K\|_{CZ_\psi} \frac{\sqrt{|I||J|}}{|K|}\phi(2^{-i})^{-d}\psi\big(2^{-i}\phi(2^{-i})^{-1}\big),$$

and the first factor is precisely the required size of the coefficients of A_K^{ij}. □

Summarizing, we have

$$\sigma_{\text{out}} = \|K\|_{CZ_\psi} \sum_{j=1}^\infty \sum_{i=j}^\infty \phi(2^{-i})^{-d}\psi\big(2^{-i}\phi(2^{-i})^{-1}\big)\langle g, S^{ij} f\rangle.$$

3.B. **Contained cubes, σ_{in}.** When $I \subsetneq J$, then I is contained in some subcube of J, which we denote by J_I.

$$\begin{aligned}\langle h_J, Th_I\rangle &= \langle 1_{J_I^c} h_J, Th_I\rangle + \langle 1_{J_I} h_J, Th_I\rangle \\ &= \langle 1_{J_I^c} h_J, Th_I\rangle + \langle h_J\rangle_{J_I}\langle 1_{J_I}, Th_I\rangle \\ &= \langle 1_{J_I^c}(h_J - \langle h_J\rangle_{J_I}), Th_I\rangle + \langle h_J\rangle_I\langle 1, Th_I\rangle,\end{aligned}$$

where we noticed that h_J is constant on $J_I \supseteq I$.

3.10. **Lemma.**

$$|\langle 1_{J_I^c}(h_J - \langle h_J\rangle_{J_I}), Th_I\rangle| \lesssim \big(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}\big)\Big(\frac{|I|}{|J|}\Big)^{1/2}\Psi\Big(\frac{\ell(I)}{\ell(J)}\phi\big(\frac{\ell(I)}{\ell(J)}\big)^{-1}\Big),$$

where

$$\Psi(r) := \int_0^r \psi(t)\frac{\mathrm{d}t}{t},$$

and $\|K\|_{CZ_0}$ could be alternatively replaced by $\|T\|_{L^2\to L^2}$.

Proof.

$$|\langle 1_{J_I^c}(h_J - \langle h_J\rangle_{J_I}), Th_I\rangle| \leq 2\|h_J\|_\infty \int_{J_I^c} |Th_I(x)|\,\mathrm{d}x,$$

where $\|h_J\|_\infty = |J|^{-1/2}$.

Case $\ell(I) \geq 2^{-r}\ell(J)$. We have

$$\begin{aligned}\int_{J_I^c} |Th_I(x)|\,\mathrm{d}x &\leq \int_{3I\setminus I}\Big|\int K(x,y)h_I(y)\,\mathrm{d}y\Big|\,\mathrm{d}x \\ &\quad + \int_{(3I)^c}\Big|\int [K(x,y) - K(x,y_I)]h_I(y)\,\mathrm{d}y\Big|\,\mathrm{d}x \\ &\lesssim \|K\|_{CZ_0}\int_{3I\setminus I}\int_I \frac{1}{|x-y|^d}\,\mathrm{d}y\,\mathrm{d}x\|h_I\|_\infty \\ &\quad + \|K\|_{CZ_\psi}\int_{(3I)^c}\frac{1}{\operatorname{dist}(x,I)^d}\psi\Big(\frac{\ell(I)}{\operatorname{dist}(x,I)}\Big)\|h_I\|_1\,\mathrm{d}x \\ &\lesssim \|K\|_{CZ_0}|I|\|h_I\|_\infty + \|K\|_{CZ_\psi}\int_{\ell(I)}^\infty \frac{1}{r^d}\psi\Big(\frac{\ell(I)}{r}\Big)r^{d-1}\,\mathrm{d}r\|h_I\|_1 \\ &= \|K\|_{CZ_0}|I|^{1/2} + \|K\|_{CZ_\psi}\int_0^1 \psi(t)\frac{\mathrm{d}t}{t}|I|^{1/2} \\ &\lesssim \big(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}\big)|I|^{1/2}\end{aligned}$$

by the Dini condition for ψ in the last step.

Alternatively, the part giving the factor $\|K\|_{CZ_0}$ could have been estimated by

$$\int_{3I\setminus I}\Big|\int K(x,y)h_I(y)\,\mathrm{d}y\Big|\,\mathrm{d}x\leq|3I\setminus I|^{1/2}\|Th_I\|_2\lesssim|I|^{1/2}\|T\|_{L^2\to L^2}.$$

Case $\ell(I)<2^{-r}\ell(J)$. Since $I\subseteq J_I$ is good, we have

$$\operatorname{dist}(I,J_I^c)>\ell(J_I)\phi\Big(\frac{\ell(I)}{\ell(J_I)}\Big)\gtrsim\ell(J)\phi\Big(\frac{\ell(I)}{\ell(J)}\Big)$$

and hence

$$\begin{aligned}\int_{J_I^c}|Th_I(x)|\,\mathrm{d}x&\lesssim\|K\|_{CZ_\psi}\int_{J_I^c}\frac{1}{d(x,I)^d}\psi\Big(\frac{\ell(I)}{\operatorname{dist}(x,I)}\Big)\|h_I\|_1\,\mathrm{d}x\\&\lesssim\|K\|_{CZ_\psi}\int_{\ell(J)\phi(\ell(I)/\ell(J))}\frac{1}{r^d}\psi\Big(\frac{\ell(I)}{r}\Big)r^{d-1}\,\mathrm{d}r\cdot\|h_I\|_1\\&=\|K\|_{CZ_\psi}\int_0^{\ell(I)/\ell(J)\cdot\phi(\ell(I)/\ell(J))^{-1}}\psi(t)\frac{\mathrm{d}t}{t}\cdot|I|^{1/2}.\end{aligned}$$ □

Now we can organize

$$\sigma'_{\mathrm{in}}:=\sum_J\sum_{I\subsetneq J}\langle g,h_J\rangle\langle 1_{J_I^c}(h_J-\langle h_J\rangle_{J_I}),Th_I\rangle\langle h_I,f\rangle=\sum_{i=1}^\infty\sum_J\sum_{\substack{I\subset J\\ \ell(I)=2^{-i}\ell(J)}},$$

and the inner sum is recognized as

$$\big(\|K\|_{CZ_0}+\|K\|_{CZ_\psi}\big)\Psi(2^{-i}\phi(2^{-i})^{-1})\langle g,A_J^{i0}f\rangle,$$

or with $\|T\|_{L^2\to L^2}$ in place of $\|K\|_{CZ_0}$, for a cancellative averaging operator of type $(i,0)$.

On the other hand,

$$\begin{aligned}\sigma''_{\mathrm{in}}&:=\sum_J\sum_{I\subsetneq J}\langle g,h_J\rangle\langle h_J\rangle_I\langle 1,Th_I\rangle\langle h_I,f\rangle\\&=\sum_I\Big\langle\sum_{J\supsetneq I}\langle g,h_J\rangle h_J\Big\rangle_I\langle 1,Th_I\rangle\langle h_I,f\rangle\\&=\sum_I\langle g\rangle_I\langle T^*1,h_I\rangle\langle h_I,f\rangle\\&=\Big\langle\sum_I\langle g\rangle_I\langle T^*1,h_I\rangle h_I,f\Big\rangle=:\langle\Pi_{T^*1}g,f\rangle=\langle g,\Pi^*_{T^*1}f\rangle.\end{aligned}$$

Here Π_{T^*1} is the *paraproduct*, a non-cancellative shift composed of the non-cancellative averaging operators

$$A_Ig=\langle T^*1,h_I\rangle\langle g\rangle_Ih_I=|I|^{-1/2}\langle T^*1,h_I\rangle\cdot\langle g,h_I^0\rangle h_I$$

of type $(0,0)$.

Summarizing, we have

$$\begin{aligned}\sigma_{\mathrm{in}}&=\sigma'_{\mathrm{in}}+\sigma''_{\mathrm{in}}\\&=\big(\|K\|_{CZ_0}+\|K\|_{CZ_\psi}\big)\sum_{i=1}^\infty\Psi(2^{-i}\phi(2^{-i})^{-1})\langle g,S^{i0}f\rangle+\langle\Pi_{T^*1}g,f\rangle,\end{aligned}$$

where $\Psi(t) = \int_0^t \psi(s)\,\mathrm{d}s/s$, and $\|K\|_{CZ_0}$ could be replaced by $\|T\|_{L^2\to L^2}$. Note that if we wanted to write Π_{T^*1} in terms of a shift with correct normalization, we should divide and multiply it by $\|T^*1\|_{\mathrm{BMO}}$, thus getting a shift times the factor $\|T^*1\|_{\mathrm{BMO}} \lesssim \|T\|_{L^2} + \|K\|_{CZ_\psi}$

3.C. **Near-by cubes, $\sigma_=$ and σ_{near}.** We are left with the sums $\sigma_=$ of equal cubes $I = J$, as well as σ_{near} of disjoint near-by cubes with $\operatorname{dist}(I, J) \le \ell(J)\phi(\ell(I)/\ell(J))$. Since I is good, this necessarily implies that $\ell(I) > 2^{-r}\ell(J)$. Then, for a given J, there are only boundedly many related I in this sum.

3.11. **Lemma.**
$$|\langle h_J, Th_I\rangle| \lesssim \|K\|_{CZ_0} + \delta_{IJ}\|T\|_{WBP}.$$

Note that if we used the L^2-boundedness of T instead of the CZ_0 and WBP conditions (as is done in Theorem 3.2, we could also estimate simply
$$|\langle h_J, Th_I\rangle| \le \|h_J\|_2\|T\|_{L^2\to L^2}\|h_I\|_2 = \|T\|_{L^2\to L^2}.$$

Proof. For disjoint cubes, we estimate directly
$$\begin{aligned}
|\langle h_J, Th_I\rangle| &\lesssim \|K\|_{CZ_0}\int_J\int_I \frac{1}{|x-y|^d}\,\mathrm{d}y\,\mathrm{d}x\|h_J\|_\infty\|h_I\|_\infty \\
&\le \|K\|_{CZ_0}\int_J\int_{3J\setminus J} \frac{1}{|x-y|^d}\,\mathrm{d}y\,\mathrm{d}x|J|^{-1/2}|I|^{-1/2} \\
&\lesssim \|K\|_{CZ_0}|J||J|^{-1/2}|J|^{-1/2} = \|K\|_{CZ_0},
\end{aligned}$$
since $|I| \eqsim |J|$.

For $J = I$, let I_i be its dyadic children. Then
$$\begin{aligned}
|\langle h_I, Th_I\rangle| &\le \sum_{i,j=1}^{2^d} |\langle h_I\rangle_{I_i}\langle h_I\rangle_{I_j}\langle 1_{I_i}, T1_{I_j}\rangle| \\
&\lesssim \|K\|_{CZ_0}\sum_{j\ne i}|I|^{-1}\int_{I_i}\int_{I_j}\frac{1}{|x-y|^d}\,\mathrm{d}x\,\mathrm{d}y + \sum_i |I|^{-1}|\langle 1_{I_i}, T1_{I_i}\rangle| \\
&\lesssim \|K\|_{CZ_0} + \|T\|_{WBP},
\end{aligned}$$
by the same estimate as earlier for the first term, and the weak boundedness property for the second. $\square$

With this lemma, the sum $\sigma_=$ is recognized as a cancellative dyadic shift of type $(0,0)$ as such:
$$\begin{aligned}
\sigma_= &= \sum_{I\in\mathscr{D}} 1_{\mathrm{good}}(I)\cdot\langle g, h_I\rangle\langle h_I, Th_I\rangle\langle h_I, f\rangle \\
&= \big(\|K\|_{CZ_0} + \|T\|_{WBP}\big)\langle g, S^{00}f\rangle,
\end{aligned}$$
where the factor in front could also be replaced by $\|T\|_{L^2\to L^2}$.

For I and J participating in σ_{near}, we conclude from Lemma 3.7 that $K := I\vee J$ satisfies $\ell(K) \le 2^r\ell(I)$, and hence we may organize
$$\sigma_{\mathrm{near}} = \sum_{i=1}^r\sum_{j=1}^i\sum_K \sum_{\substack{I,J:I\vee J=K \\ \operatorname{dist}(I,J)\le\ell(J)\phi(\ell(I)/\ell(J)) \\ \ell(I)=2^{-i}\ell(K) \\ \ell(J)=2^{-j}\ell(K)}} ,$$

and the innermost sum is recognized as $\|K\|_{CZ_0}\langle g, A_K^{ij} f\rangle$ for some cancellative averaging operator of type (i,j).

Summarizing, we have

$$\sigma_{=} + \sigma_{\text{near}} = \big(\|K\|_{CZ_0} + \|T\|_{WBP}\big)\langle g, S^{00} f\rangle + \|K\|_{CZ_0}\sum_{j=1}^{r}\sum_{i=j}^{r}\langle g, S^{ij} f\rangle,$$

where S^{00} and S^{ij} are cancellative dyadic shifts, and the factor $\big(\|K\|_{CZ_0}+\|T\|_{WBP}\big)$ could also be replaced by $\|T\|_{L^2\to L^2}$.

3.D. **Synthesis.** We have checked that

$$\begin{aligned}
&\sum_{\ell(I)\leq \ell(J)} 1_{\text{good}}(I)\langle g, h_J\rangle\langle h_J, T h_I\rangle\langle h_I, f\rangle \\
&\qquad = \big(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}\big)\Big(\sum_{1\leq j\leq i<\infty} \phi(2^{-i})^{-d}\psi(2^{-i}\phi(2^{-i})^{-1})\langle g, S^{ij} f\rangle \\
&\qquad\qquad + \sum_{1\leq i<\infty}\Psi(2^{-i}\phi(2^{-i})^{-1}))\langle g, S^{i0} f\rangle\Big) \\
&\qquad\quad + \big(\|K\|_{CZ_0} + \|T\|_{WBP}\big)\langle g, S^{00} f\rangle + \langle g, \Pi_{T^*1}^{*} f\rangle
\end{aligned}$$

where $\Psi(t) = \int_0^t \psi(s)\,\mathrm{d}s/s$, Π_{T^*1} is a paraproduct—a non-cancellative shift of type $(0,0)$–, and all other S^{ij} is a cancellative dyadic shifts of type (i,j).

By symmetry (just observing that the cubes of equal size contributed precisely to the presence of the cancellative shifts of type (i,i), and that the dual of a shift of type (i,j) is a shift of type (j,i)), it follows that

$$\begin{aligned}
&\sum_{\ell(I)>\ell(J)} 1_{\text{good}}(J)\langle g, h_J\rangle\langle h_J, T h_I\rangle\langle h_I, f\rangle \\
&\qquad = \big(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}\big)\Big(\sum_{1\leq i<j<\infty} \phi(2^{-j})^{-d}\psi(2^{-j}\phi(2^{-j})^{-1})\langle g, S^{ij} f\rangle \\
&\qquad\quad + \sum_{1\leq j<\infty}\Psi(2^{-j}\phi(2^{-j})^{-1}))\langle g, S^{0j} f\rangle\Big) + \langle g, \Pi_{T1} f\rangle
\end{aligned}$$

so that altogether

$$\begin{aligned}
&\sum_{I,J} 1_{\text{good}}(\min\{I,J\})\langle g, h_J\rangle\langle h_J, T h_I\rangle\langle h_I, f\rangle \\
&\quad = \big(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}\big)\Big(\sum_{i=1}^{\infty}\Psi(2^{-i}\phi(2^{-i})^{-1}))(\langle g, S^{i0} f\rangle + \langle g, S^{0i} f\rangle) \\
&\qquad + \sum_{i,j=1}^{\infty}\phi(2^{-\max(i,j)})^{-d}\psi(2^{-\max(i,j)}\phi(2^{-\max(i,j)})^{-1})\langle g, S^{ij} f\rangle\Big) \\
&\qquad + \big(\|K\|_{CZ_0} + \|T\|_{WBP}\big)\langle g, S^{00} f\rangle + \langle g, \Pi_{T1} f\rangle + \langle g, \Pi_{T^*1}^{*} f\rangle,
\end{aligned}$$

and this completes the proof of Theorem 3.2.

4. Two-weight theory for dyadic shifts

Before proceeding further, it is convenient to introduce a useful trick due to E. Sawyer. Let σ be an everywhere positive, finitely-valued function. Then $f \in$

$L^p(w)$ if and only if $\phi = f/\sigma \in L^p(\sigma^p w)$, and they have equal norms in the respective spaces. Hence an inequality

$$\|Tf\|_{L^p(w)} \leq N\|f\|_{L^p(w)} \qquad \forall f \in L^p(w) \tag{4.1}$$

is equivalent to

$$\|T(\phi\sigma)\|_{L^p(w)} \leq N\|\phi\sigma\|_{L^p(w)} = N\|\phi\|_{L^p(\sigma^p w)} \qquad \forall \phi \in L^p(\sigma^p w).$$

This is true for any σ, and we now choose it in such a way that $\sigma^p w = \sigma$, i.e., $\sigma = w^{-1/(p-1)} = w^{1-p'}$, where p' is the dual exponent. So finally (4.1) is equivalent to

$$\|T(\phi\sigma)\|_{L^p(w)} \leq N\|\phi\|_{L^p(\sigma)} \qquad \forall \phi \in L^p(\sigma).$$

This formulation has the advantage that the norm on the right and the operator

$$T(\phi\sigma)(x) = \int K(x,y)\phi(y) \cdot \sigma(y)\,\mathrm{d}y$$

involve integration with respect to the same measure σ. In particular, the A_2 theorem is equivalent to

$$\|T(f\sigma)\|_{L^2(w)} \leq c_T[w]_{A_2}\|f\|_{L^2(\sigma)}$$

for all $f \in L^2(w)$, for all $w \in A_2$ and $\sigma = w^{-1}$. But once we know this, we can also study this two-weight inequality on its own right, for two general measures w and σ, which need not be related by the pointwise relation $\sigma(x) = 1/w(x)$.

4.2. **Theorem.** *Let σ and w be two locally finite measures with*

$$[w,\sigma]_{A_2} := \sup_Q \frac{w(Q)\sigma(Q)}{|Q|^2} < \infty.$$

Then a dyadic shift S of type (i,j) satisfies $S(\sigma\cdot) : L^2(\sigma) \to L^2(w)$ if and only if

$$\mathfrak{S} := \sup_Q \frac{\|1_Q S(\sigma 1_Q)\|_{L^2(w)}}{\sigma(Q)^{1/2}}, \qquad \mathfrak{S}^* := \sup_Q \frac{\|1_Q S^*(w 1_Q)\|_{L^2(\sigma)}}{w(Q)^{1/2}}$$

are finite, and in this case

$$\|S(\sigma\cdot)\|_{L^2(\sigma)\to L^2(w)} \lesssim (1+\kappa)(\mathfrak{S}+\mathfrak{S}^*) + (1+\kappa)^2[w,\sigma]_{A_2}^{1/2},$$

where $\kappa = \max\{i,j\}$.

This result from my work with Pérez, Treil, and Volberg [21] was preceded by an analogous qualitative version due to Nazarov, Treil, and Volberg [34].

The proof depends on decomposing functions in the spaces $L^2(w)$ and $L^2(\sigma)$ in terms of expansions similar to the Haar expansion in $L^2(\mathbb{R}^d)$. Let $\mathbb{D}_I^\sigma$ be the orthogonal projection of $L^2(\sigma)$ onto its subspace of functions supported on I, constant on the subcubes of I, and with vanishing integral with respect to $\mathrm{d}\sigma$. Then any two $\mathbb{D}_I^\sigma$ are orthogonal to each other. Under the additional assumption that the σ measure of quadrants of $\mathbb{R}^d$ is infinite, we have the expansion

$$f = \sum_{Q\in\mathscr{D}} \mathbb{D}_Q^\sigma f$$

for all $f \in L^2(\sigma)$, and Pythagoras' theorem says that

$$\|f\|_{L^2(\sigma)} = \Big(\sum_{Q\in\mathscr{D}} \|\mathbb{D}_Q^\sigma f\|_{L^2(\sigma)}^2\Big)^{1/2}.$$

(These formulae needs a slight adjustment if the σ measure of quadrants is finite; Theorem 4.2 remains true without this extra assumption.) Let us also write

$$\mathbb{D}_K^{\sigma,i} := \sum_{\substack{I\subseteq K\\ \ell(I)=2^{-i}\ell(K)}} \mathbb{D}_I^\sigma.$$

For a fixed $i\in\mathbb{N}$, these are also orthogonal to each other, and the above formulae generalize to

$$f=\sum_{Q\in\mathscr{D}}\mathbb{D}_Q^{\sigma,i}f,\qquad \|f\|_{L^2(\sigma)}=\Big(\sum_{Q\in\mathscr{D}}\|\mathbb{D}_Q^{\sigma,i}f\|_{L^2(\sigma)}^2\Big)^{1/2}.$$

The proof is in fact very similar in spirit to that of Theorem 3.2; it is another $T1$ argument, but now with respect to the measures σ and w in place of the Lebesgue measure. We hence expand

$$\langle g,S(\sigma f)\rangle_w=\sum_{Q,R\in\mathscr{D}}\langle\mathbb{D}_R^w g,S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w,\qquad f\in L^2(\sigma),\ g\in L^2(w),$$

and estimate the matrix coefficients

$$\begin{aligned}\langle\mathbb{D}_R^w g,S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w&=\sum_K\langle\mathbb{D}_R^w g,A_K(\sigma\mathbb{D}_Q^\sigma f)\rangle_w\\ &=\sum_K\sum_{I,J\subseteq K}a_{IJK}\langle\mathbb{D}_R^w g,h_J\rangle_w\langle h_I,\mathbb{D}_Q^\sigma f\rangle_\sigma.\end{aligned}\tag{4.3}$$

For $\langle h_I,\mathbb{D}_Q^\sigma f\rangle_\sigma\neq 0$, there must hold $I\cap Q\neq\varnothing$, thus $I\subseteq Q$ or $Q\subsetneq I$. But in the latter case h_I is constant on Q, while $\int\mathbb{D}_Q^\sigma f\cdot\sigma=0$, so the pairing vanishes even in this case. Thus the only nonzero contributions come from $I\subseteq Q$, and similarly from $J\subseteq R$. Since $I,J\subseteq K$, there holds

$$\big(I\subseteq Q\subsetneq K\quad\text{or}\quad K\subseteq Q\big)\qquad\text{and}\qquad\big(J\subseteq R\subsetneq K\quad\text{or}\quad K\subseteq R\big).$$

4.A. **Disjoint cubes.** Suppose now that $Q\cap R=\varnothing$, and let K be among those cubes for which A_K gives a nontrivial contribution above. Then it cannot be that $K\subseteq Q$, since this would imply that $Q\cap R\supseteq K\cap J=J\neq\varnothing$, and similarly it cannot be that $K\subseteq R$. Thus $Q,R\subsetneq K$, and hence

$$Q\vee R\subseteq K.$$

Then

$$\begin{aligned}|\langle\mathbb{D}_R^w g,S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w|&\leq\sum_{K\supseteq Q\vee R}|\langle\mathbb{D}_R^w g,A_K(\sigma\mathbb{D}_Q^\sigma f)\rangle_w|\\ &\lesssim\sum_{K\supseteq Q\vee R}\frac{\|\mathbb{D}_R^w g\|_{L^1(w)}\|\mathbb{D}_Q^\sigma f\|_{L^1(\sigma)}}{|K|}\\ &\lesssim\frac{\|\mathbb{D}_R^w g\|_{L^1(w)}\|\mathbb{D}_Q^\sigma f\|_{L^1(\sigma)}}{|Q\vee R|}\end{aligned}$$

On the other hand, we have $Q\supseteq I$, $R\supseteq J$ for some $I,J\subseteq K$ with $\ell(I)=2^{-i}\ell(K)$ and $\ell(J)=2^{-j}\ell(K)$. Hence $2^{-i}\ell(K)\leq\ell(Q)$ and $2^{-j}\ell(K)\leq\ell(R)$, and thus

$$Q\vee R\subseteq K\subseteq Q^{(i)}\cap R^{(j)}.$$

Now it is possible to estimate the total contribution of the part of the matrix with $Q\cap R=\varnothing$. Let $P:=Q\vee R$ be a new auxiliary summation variable. Then

$Q, R \subset P$, and $\ell(Q) = 2^{-a}\ell(P)$, $\ell(R) = 2^{-b}\ell(P)$ where $a = 1, \dots, i$, $b = 1, \dots, j$. Thus

$$\begin{aligned}
&\sum_{\substack{Q,R\in\mathscr{D}\\ Q\cap R=\varnothing}} |\langle \mathbb{D}_R^w g, S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w| \\
&\qquad\lesssim \sum_{a=1}^{i}\sum_{b=1}^{j}\sum_{P\in\mathscr{D}} \frac{1}{|P|} \sum_{\substack{Q,R\in\mathscr{D}:Q\vee R=P\\ \ell(Q)=2^{-a}\ell(P)\\ \ell(R)=2^{-b}\ell(P)}} \|\mathbb{D}_R^w g\|_{L^1(\sigma)} \|\mathbb{D}_Q^\sigma f\|_{L^1(w)} \\
&\qquad\le \sum_{a,b=1}^{i,j}\sum_{P\in\mathscr{D}} \frac{1}{|P|} \sum_{\substack{R\subseteq P\\ \ell(R)=2^{-b}\ell(P)}} \|\mathbb{D}_R^w g\|_{L^1(\sigma)} \sum_{\substack{Q\subseteq P\\ \ell(Q)=2^{-a}\ell(P)}} \|\mathbb{D}_Q^\sigma f\|_{L^1(\sigma)} \\
&\qquad= \sum_{a,b=1}^{i,j}\sum_{P\in\mathscr{D}} \frac{1}{|P|} \Big\| \sum_{\substack{R\subseteq P\\ \ell(R)=2^{-b}\ell(P)}} \mathbb{D}_R^w g \Big\|_{L^1(\sigma)} \Big\| \sum_{\substack{Q\subseteq P\\ \ell(Q)=2^{-a}\ell(P)}} \mathbb{D}_Q^\sigma f \Big\|_{L^1(\sigma)} \\
&\qquad\qquad\qquad \text{(by disjoint supports)} \\
&\qquad= \sum_{a,b=1}^{i,j}\sum_{P\in\mathscr{D}} \frac{1}{|P|} \|\mathbb{D}_P^{w,j} g\|_{L^1(w)} \|\mathbb{D}_P^{\sigma,i} f\|_{L^1(\sigma)} \\
&\qquad\le \sum_{a,b=1}^{i,j}\sum_{P\in\mathscr{D}} \frac{\sigma(P)^{1/2} w(P)^{1/2}}{|P|} \|\mathbb{D}_P^{w,j} g\|_{L^2(w)} \|\mathbb{D}_P^{\sigma,i} f\|_{L^2(\sigma)} \\
&\qquad\le \sum_{a,b=1}^{i,j} [w,\sigma]_{A_2}^{1/2} \Big(\sum_{P\in\mathscr{D}} \|\mathbb{D}_P^{w,j} g\|_{L^2(w)}^2\Big)^{1/2} \Big(\sum_{P\in\mathscr{D}} \|\mathbb{D}_P^{\sigma,i} f\|_{L^2(\sigma)}^2\Big)^{1/2} \\
&\qquad\le ij[w,\sigma]_{A_2}^{1/2} \|g\|_{L^2(w)} \|f\|_{L^2(\sigma)}.
\end{aligned}$$

4.B. **Deeply contained cubes.** Consider now the part of the sum with $Q \subset R$ and $\ell(Q) < 2^{-i}\ell(R)$. (The part with $R \subset Q$ and $\ell(R) < 2^{-j}\ell(Q)$ would be handled in a symmetrical manner.)

4.4. **Lemma.** *For all $Q \subset R$ with $\ell(Q) < 2^{-i}\ell(R)$, we have*

$$\langle \mathbb{D}_R^w g, S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w = \langle \mathbb{D}_R^w g\rangle_{Q^{(i)}} \langle S^*(w1_{Q^{(i)}}), \mathbb{D}_Q^\sigma f\rangle_\sigma,$$

where further

$$\mathbb{D}_Q^\sigma S^*(w1_{Q^{(i)}}) = \mathbb{D}_Q^\sigma S^*(w1_P) \qquad \textit{for any } P \supseteq Q^{(i)}.$$

Recall that $\mathbb{D}_Q^\sigma = (\mathbb{D}_Q^\sigma)^2 = (\mathbb{D}_Q^\sigma)^*$ is an orthogonal projection on $L^2(\sigma)$, so that it can be moved to either or both sides of the pairing $\langle\ ,\ \rangle_\sigma$.

Proof. Recall formula (4.3). If $\langle h_I, \mathbb{D}_Q^\sigma f\rangle_\sigma$ is nonzero, then $I \subseteq Q$, and hence

$$J \subseteq K = I^{(i)} \subseteq Q^{(i)} \subsetneq R$$

for all J participating in the same A_K as I. Thus $\mathbb{D}^w_R g$ is constant on $Q^{(i)}$, hence

$$\begin{aligned}\langle \mathbb{D}^w_R g, A_K(\sigma\mathbb{D}^\sigma_Q f)\rangle_w &= \langle 1_{Q^{(i)}}\mathbb{D}^w_R g, A_K(\sigma\mathbb{D}^\sigma_Q f)\rangle_w \\ &= \langle \mathbb{D}^w_R g\rangle^w_{Q^{(i)}}\langle 1_{Q^{(i)}}, A_K(\sigma\mathbb{D}^\sigma_Q f)\rangle_w \\ &= \langle \mathbb{D}^w_R g\rangle^w_{Q^{(i)}}\langle A^*_K(w1_{Q^{(i)}}), \mathbb{D}^\sigma_Q f\rangle_\sigma.\end{aligned}$$

Moreover, for any $P \supseteq Q^{(i)} \supseteq K$,

$$\begin{aligned}\langle \mathbb{D}^\sigma_Q A^*_K(w1_{Q^{(i)}}), f\rangle_\sigma &= \langle 1_{Q^{(i)}}, A_K(\sigma\mathbb{D}^\sigma_Q f)\rangle_w \\ &= \int A_K(\sigma\mathbb{D}^\sigma_Q f)w \\ &= \langle 1_P, A_K(\sigma\mathbb{D}^\sigma_Q f)\rangle_w = \langle \mathbb{D}^\sigma_Q A^*_K(w1_P), f\rangle_\sigma.\end{aligned}$$

Summing these equalities over all relevant K, and using $S = \sum_K A_K$, gives the claim. $\square$

By the lemma, we can then manipulate

$$\begin{aligned}&\sum_{\substack{Q,R:Q\subset R\\ \ell(Q)<2^{-i}\ell(R)}}\langle \mathbb{D}^w_R g, S(\sigma\mathbb{D}^\sigma_Q f)\rangle_w \\ &\qquad= \sum_Q\Big(\sum_{R\supsetneq Q^{(i)}}\langle \mathbb{D}^w_R g\rangle^w_{Q^{(i)}}\Big)\langle S^*(w1_{Q^{(i)}}), \mathbb{D}^\sigma_Q f\rangle_\sigma \\ &\qquad= \sum_Q \langle g\rangle^w_{Q^{(i)}}\langle S^*(w1_{Q^{(i)}}), \mathbb{D}^\sigma_Q f\rangle_\sigma \\ &\qquad= \sum_R \langle g\rangle^w_R\Big\langle S^*(w1_R), \sum_{\substack{Q\subseteq R\\ \ell(Q)=2^{-i}\ell(R)}}\mathbb{D}^\sigma_Q f\Big\rangle_\sigma \\ &\qquad= \sum_R \langle g\rangle^w_R\Big\langle S^*(w1_R), \mathbb{D}^{\sigma,i}_R f\Big\rangle_\sigma,\end{aligned}$$

where $\langle g\rangle^w_R := w(R)^{-1}\int_R g\cdot w$ is the average of g on R with respect to the w measure.

By using the properties of the pairwise orthogonal projections $\mathbb{D}^{\sigma,i}_R$ on $L^2(\sigma)$, the above series may be estimated as follows:

$$\begin{aligned}&\Big|\sum_{\substack{Q,R:Q\subset R\\ \ell(Q)<2^{-i}\ell(R)}}\langle \mathbb{D}^w_R g, S(\sigma\mathbb{D}^\sigma_Q f)\rangle_w\Big| \\ &\leq \sum_R |\langle g\rangle^w_R|\|\mathbb{D}^{\sigma,i}_R S^*(w1_R)\|_{L^2(\sigma)}\|\mathbb{D}^{\sigma,i}_R f\|_{L^2(\sigma)} \\ &\leq \Big(\sum_R |\langle g\rangle^w_R|^2\|\mathbb{D}^{\sigma,i}_R S^*(w1_R)\|^2_{L^2(\sigma)}\Big)^{1/2}\Big(\sum_R \|\mathbb{D}^{\sigma,i}_R f\|^2_{L^2(\sigma)}\Big)^{1/2},\end{aligned}$$

where the last factor is equal to $\|f\|_{L^2(w)}$.

The first factor on the right is handled by the dyadic Carleson embedding theorem: It follows from the second equality of Lemma 4.4, namely $\mathbb{D}^\sigma_Q S^*(w1^{(i)}_Q) = \mathbb{D}^\sigma_Q S^*(w1_P)$ for all $P \supseteq Q^{(i)}$, that $\mathbb{D}^{\sigma,i}_R S^*(w1_R) = \mathbb{D}^\sigma_Q S^*(w1_P)$ for all $P \subseteq R$.

Hence, we have

$$\begin{aligned}\sum_{R\subseteq P}\|\mathbb{D}_R^{\sigma,i}S^*(w1_R)\|_{L^2(\sigma)}^2&=\sum_{R\subseteq P}\|\mathbb{D}_R^{\sigma,i}(1_PS^*(w1_P))\|_{L^2(\sigma)}^2\\&\leq\|1_PS^*(w1_P)\|_{L^2(\sigma)}^2\lesssim\mathfrak{S}_*^2\sigma(P)\end{aligned}$$

by the (dual) testing estimate for the dyadic shifts. By the Carleson embedding theorem, it then follows that

$$\Big(\sum_R|\langle g\rangle_R^w|^2\|\mathbb{D}_R^{\sigma,i}S^*(w1_R)\|_{L^2(\sigma)}^2\Big)^{1/2}\lesssim\mathfrak{S}_*\|g\|_{L^2(\sigma)},$$

and the estimation of the deeply contained cubes is finished.

4.C. **Contained cubes of comparable size.** It remains to estimate

$$\sum_{\substack{Q,R:Q\subseteq R\\ \ell(Q)\geq 2^{-i}\ell(R)}}\langle\mathbb{D}_R^wg,S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w;$$

the sum over $R\subsetneq Q$ with $\ell(R)\geq 2^{-j}\ell(Q)$ would be handled in a symmetric manner. The sum of interest may be written as

$$\sum_{a=0}^i\sum_R\sum_{\substack{Q\subseteq R\\ \ell(Q)=2^{-a}\ell(R)}}\langle\mathbb{D}_R^wg,S(\sigma\mathbb{D}_Q^\sigma f)\rangle_w=\sum_{a=0}^i\sum_R\langle\mathbb{D}_R^wg,S(\sigma\mathbb{D}_R^{\sigma,i}f)\rangle_w,$$

and

$$\langle\mathbb{D}_R^wg,S(\sigma\mathbb{D}_R^{\sigma,i}f)\rangle_w=\sum_{k=1}^{2^d}\langle\mathbb{D}_R^wg\rangle_{R_k}\langle S^*(w1_{R_k}),\mathbb{D}_R^{\sigma,i}f\rangle_\sigma$$

where the R_k are the 2^d dyadic children of R, and $\langle\mathbb{D}_R^wg\rangle_{R_k}$ is the constant valued of $\mathbb{D}_R^wg$ on R_k. Now

$$\langle S^*(w1_{R_k}),\mathbb{D}_R^{\sigma,i}f\rangle_\sigma=\langle 1_{R_k}S^*(w1_{R_k}),\mathbb{D}_R^{\sigma,i}f\rangle_\sigma+\langle S^*(w1_{R_k}),1_{R_k^c}\mathbb{D}_R^{\sigma,i}f\rangle_\sigma,$$

where

$$|\langle 1_{R_k}S^*(w1_{R_k}),\mathbb{D}_R^{\sigma,i}f\rangle_\sigma|\leq\mathfrak{S}_*w(R_k)^{1/2}\|\mathbb{D}_R^{\sigma,i}f\|_{L^2(\sigma)}$$

and, observing that only those A_K^* where K intersects both R_k and R_k^c contribute to the second part,

$$\begin{aligned}|\langle S^*(w1_{R_k}),1_{R_k^c}\mathbb{D}_R^{\sigma,i}f\rangle_\sigma|&=\Big|\sum_{K\supsetneq R_k}\langle A_K^*(w1_{R_k}),1_{R_k^c}\mathbb{D}_R^{\sigma,i}f\rangle_\sigma\Big|\\&\lesssim\sum_{K\supseteq R}\frac{1}{|K|}w(R_k)\|\mathbb{D}_R^{\sigma,i}f\|_{L^1(\sigma)}\\&\lesssim\frac{1}{|R|}w(R_k)\sigma(R)^{1/2}\|\mathbb{D}_R^{\sigma,i}f\|_{L^1(\sigma)}\\&\leq\frac{w(R)^{1/2}\sigma(R)^{1/2}}{|R|}w(R_k)^{1/2}\|\mathbb{D}_R^{\sigma,i}f\|_{L^2(\sigma)}\\&\leq[w,\sigma]_{A_2}w(R_k)^{1/2}\|\mathbb{D}_R^{\sigma,i}f\|_{L^2(\sigma)}.\end{aligned}$$

It follows that

$$|\langle S^*(w1_{R_k}),\mathbb{D}_R^{\sigma,i}f\rangle_\sigma|\lesssim(\mathfrak{S}_*+[w,\sigma]_{A_2})w(R_k)^{1/2}\|\mathbb{D}_R^{\sigma,i}f\|_{L^2(\sigma)}$$

and hence

$$|\langle \mathbb{D}^w_R g, S(\sigma \mathbb{D}^{\sigma,i}_R f)\rangle_w| \lesssim (\mathfrak{S}_* + [w,\sigma]_{A_2})\|\mathbb{D}^w_R g\|_{L^2(w)}\|\mathbb{D}^{\sigma,i}_R f\|_{L^2(\sigma)}$$

Finally,

$$\begin{aligned}
&\sum_{a=0}^{i}\sum_R |\langle \mathbb{D}^w_R g, S(\sigma \mathbb{D}^{\sigma,i}_R f)\rangle_w| \\
&\lesssim (\mathfrak{S}_* + [w,\sigma]_{A_2})\sum_{a=0}^{i}\Big(\sum_R \|\mathbb{D}^w_R g\|^2_{L^2(\sigma)}\Big)^{1/2}\Big(\sum_R \|\mathbb{D}^{\sigma,i}_R f\|^2_{L^2(\sigma)}\Big)^{1/2} \\
&\leq (1+i)(\mathfrak{S}_* + [w,\sigma]_{A_2})\|g\|_{L^2(w)}\|f\|_{L^2(\sigma)}.
\end{aligned}$$

The symmetric case with $R \subset Q$ with $\ell(R) \geq 2^{-j}\ell(Q)$ similarly yields the factor $(1+j)(\mathfrak{S} + [w,\sigma]_{A_2})$. This completes the proof of Theorem 4.2.

5. Final decompositions: verification of the testing conditions

We now turn to the estimation of the testing constant

$$\mathfrak{S} := \sup_{Q\in\mathscr{D}} \frac{\|1_Q S(\sigma 1_Q)\|_{L^2(w)}}{\sigma(Q)^{1/2}}.$$

Bounding $\mathfrak{S}_*$ is analogous by exchanging the roles of w and σ.

5.A. **Several splittings.** First observe that

$$1_Q S(\sigma 1_Q) = 1_Q \sum_{K:K\cap Q\neq\varnothing} A_K(\sigma 1_Q) = 1_Q \sum_{K\subseteq Q} A_K(\sigma 1_Q) + 1_Q \sum_{K\supsetneq Q} A_K(\sigma 1_Q).$$

The second part is immediate to estimate even pointwise by

$$|1_Q A_K(\sigma 1_Q)| \leq 1_Q \frac{\sigma(Q)}{|K|}, \qquad \sum_{K\supsetneq Q}\frac{1}{|K|} \leq \frac{1}{|Q|},$$

and hence its $L^2(w)$ norm is bounded by

$$\Big\|1_Q\frac{\sigma(Q)}{|Q|}\Big\|_{L^2(w)} = \frac{w(Q)^{1/2}\sigma(Q)}{|Q|} \leq [w,\sigma]_{A_2}\sigma(Q)^{1/2}.$$

So it remains to concentrate on $K\subseteq Q$, and we perform several consecutive splittings of this collection of cubes. First, we **separate scales** by introducing the splitting according to the $\kappa+1$ possible values of $\log_2\ell(K) \mod (\kappa+1)$. We denote a generic choice of such a collection by

$$\mathscr{K} = \mathscr{K}_k := \{K\subseteq Q : \log_2\ell(K) \equiv k \mod (\kappa+1)\},$$

where k is arbitrary but fixed. (We will drop the subscript k, since its value plays no role in the subsequent argument.) Next, we **freeze the A_2 characteristic** by setting

$$\mathscr{K}^a := \Big\{K\in\mathscr{K} : 2^{a-1} < \frac{w(K)\sigma(K)}{|K|} \leq 2^a\Big\}, \qquad a\in\mathbb{Z}, \quad a \leq \lceil \log_2[w,\sigma]_{A_2}\rceil,$$

where $\lceil\ \rceil$ means rounding up to the next integer.

In the next step, we **choose the principal cubes** $P\in\mathscr{P}^a\subseteq\mathscr{K}^a$. This construction was first introduced by B. Muckenhoupt and R. Wheeden [31], and it has been influential ever since. Let $\mathscr{P}_0^a$ consist of all maximal cubes in $\mathscr{K}^a$, and inductively $\mathscr{P}_{p+1}^a$ consist of all maximal $P' \in \mathscr{K}^a$ such that

$$P' \subset P \in \mathscr{P}_p^a, \qquad \frac{\sigma(P')}{|P'|} > 2\frac{\sigma(P)}{|P|}.$$

Finally, let $\mathscr{P}^a := \bigcup_{p=0}^\infty \mathscr{P}_p^a$. For each $K \in \mathscr{K}^a$, let $\Pi^a(K)$ denote the minimal $P \in \mathscr{P}^a$ such that $K \subseteq P$. Then we set

$$\mathscr{K}^a(P) := \{K \in \mathscr{K}^a : \Pi^a(K) = P\}, \qquad P \in \mathscr{P}^a.$$

Note that $\sigma(K)/|K| \leq 2\sigma(P)/|P|$ for all $K \in \mathscr{K}^a(P)$, which allows us to **freeze the σ-to-Lebesgue measure ratio** by the final subcollections

$$\mathscr{K}_b^a(P) := \Big\{K \in \mathscr{K}^a(P) : 2^{-b} < \frac{\sigma(K)}{|K|}\frac{|P|}{\sigma(P)} \leq 2^{1-b}\Big\}, \qquad b \in \mathbb{N}.$$

We have

$$\{K \in \mathscr{D} : K \subseteq Q\} = \bigcup_{k=0}^{\kappa} \mathscr{K}_k, \qquad \mathscr{K}_k = \mathscr{K} = \bigcup_{a \leq \lceil \log_2 [w,\sigma]_{A_2} \rceil} \mathscr{K}^a,$$

$$\mathscr{K}^a = \bigcup_{P \in \mathscr{P}^a} \mathscr{K}^a(P), \qquad \mathscr{K}^a(P) = \bigcup_{b=0}^{\infty} \mathscr{K}_b^a(P),$$

where all unions are disjoint. Note that we drop the reference to the separation-of-scales parameter k, since this plays no role in the forthcoming arguments. Recalling the notation for subshifts $S_{\mathscr{Q}} = \sum_{K \in \mathscr{Q}} A_K$, this splitting of collections of cubes leads to the splitting of the function

$$\sum_{K \subseteq Q} A_K(\sigma 1_Q) = \sum_{k=0}^{\kappa} \sum_{a \leq \lceil \log_2 [w,\sigma]_{A_2} \rceil} \sum_{P \in \mathscr{P}^a} \sum_{b=0}^{\infty} S_{\mathscr{K}_b^a(P)}(\sigma 1_Q).$$

On the level of the function, we split one more time to write

$$\begin{aligned} S_{\mathscr{K}_b^a(P)}(\sigma 1_Q) &= \sum_{n=0}^{\infty} 1_{E_b^a(P,n)} S_{\mathscr{K}_b^a(P)}(\sigma 1_Q), \\ E_b^a(P,n) &:= \{x \in \mathbb{R}^d : n2^{-b}\langle\sigma\rangle_P < |S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)(x)| \leq (n+1)2^{-b}\langle\sigma\rangle_P\}. \end{aligned}$$

This final splitting, from [18], is not strictly 'necessary' in that it was not part of the original argument in [15], nor its predecessor in [24], which made instead more careful use of the cubes where $S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)$ stays constant; however, it now seems that this splitting provides another simplification of the argument.

Now all relevant cancellation is inside the functions $S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)$, so that we can simply estimate by the triangle inequality:

$$\begin{aligned} &\Big| \sum_{K \subseteq Q} A_K(\sigma 1_Q) \Big| \\ &\quad \leq \sum_{k=0}^{\kappa} \sum_{a} \sum_{P \in \mathscr{P}^a} \sum_{b=0}^{\infty} \sum_{n=0}^{\infty} (1+n)2^{-b}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b}\langle\sigma\rangle_P\}}, \end{aligned}$$

and

$$\begin{aligned}\Big\|&\sum_{K\subseteq Q}A_K(\sigma 1_Q)\Big\|_{L^2(w)}\\ &\leq\sum_{k=0}^{\kappa}\sum_a\sum_{b=0}^\infty 2^{-b}\sum_{n=0}^\infty(1+n)\Big\|\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}\Big\|_{L^2(w)}.\end{aligned}$$

Obviously, we will need good estimates to be able to sum up these infinite series.

Write the last norm as

$$\Big(\int\Big[\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}(x)\Big]^2\,\mathrm{d}w(x)\Big)^{1/2},$$

observe that

$$\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}\subseteq P,$$

and look at the integrand at a fixed point $x\in\mathbb{R}^d$. At this point we sum over a subset of those values of $\langle\sigma\rangle_P$ where the principal cube $P\ni x$. Let P_0 be the smallest cube such that $|S_{\mathscr{K}_b^a(P)}|>n2^{-b}\langle\sigma\rangle_P$, let P_1 be the next smallest, and so on. Then $\langle\sigma\rangle_{P_m}<2^{-1}\langle\sigma\rangle_{P_{m-1}}<\ldots<2^{-m}\langle\sigma\rangle_{P_0}$ by the construction of the principal cubes, and hence

$$\begin{aligned}\Big[\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}|>n2^{-b}\langle\sigma\rangle_P\}}(x)\Big]^2&=\Big[\sum_{m=0}^\infty\langle\sigma\rangle_{P_m}\Big]^2\\ &\leq\Big[\sum_{m=0}^\infty 2^{-m}\langle\sigma\rangle_{P_0}\Big]^2=4\langle\sigma\rangle_{P_0}^2\\ &\leq 4\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P^2 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}(x)\end{aligned}$$

Hence

$$\begin{aligned}\Big\|&\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}\Big\|_{L^2(w)}\\ &\leq\Big(\int\Big[4\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P^2 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}\Big]w\Big)^{1/2}\\ &=2\Big(\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P^2 w(\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\})\Big)^{1/2},\end{aligned}$$

and it remains to obtain good estimates for the measure of the level sets

$$\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}.$$

5.B. **Weak-type and John–Nirenberg-style estimates.** We still need to estimate the sets above. Recall that $S_{\mathscr{K}_b^a(P)}$ is a subshift of S, which in particular has its scales separated so that $\log_2\ell(K)\equiv k \mod (\kappa+1)$ for all K for which A_K participating in $S_{\mathscr{K}_b^a(P)}$ is nonzero and $k\in\{0,1,\ldots,\kappa:=\max\{i,j\}\}$ is fixed, S being of type (i,j). The following estimate deals with such subshifts, which we simply denote by S.

5.1. **Proposition.** *Let S be a dyadic shift of type (i,j) with scales separated. Then*

$$|\{|Sf|>\lambda\}|\leq\frac{C}{\lambda}\|f\|_1,\qquad\forall\lambda>0,$$

where C depends only on the dimension.

Proof. The proof uses the classical Calderón–Zygmund decomposition:

$$f = g + b, \qquad b := \sum_{L \in \mathscr{B}} b_L := \sum_{L \in \mathscr{B}} 1_B(f - \langle f \rangle_L),$$

where $L \in \mathscr{B}$ are the maximal dyadic cubes with $\langle |f| \rangle_L > \lambda$; hence $\langle |f| \rangle_L \le 2^d \lambda$. As usual,

$$g = f - b = 1_{\left(\bigcup \mathscr{B}\right)^c} f + \sum_{L \in \mathscr{B}} \langle f \rangle_L$$

satisfies $\|g\|_\infty \le 2^d \lambda$ and $\|g\|_1 \le \|f\|_1$, hence $\|g\|_2^2 \le \|g\|_\infty \|g\|_1 \le 2^d \lambda \|f\|_1$, and thus

$$|\{|Sg| > \tfrac{1}{2}\lambda\}| \le \frac{4}{\lambda^2}\|Sg\|_2^2 \le \frac{4}{\lambda^2}\|g\|_2^2 \le 4 \cdot 2^d \frac{1}{\lambda}\|f\|_1.$$

It remains to estimate $\{|Sb| > \frac{1}{2}\lambda\}$. First observe that

$$Sb = \sum_{K \in \mathscr{D}} \sum_{L \in \mathscr{B}} A_K b_L = \sum_{L \in \mathscr{B}} \Big(\sum_{K \subseteq L} A_K b_L + \sum_{K \supsetneq L} A_K b_L \Big),$$

since $A_K b_L \ne 0$ only if $K \cap L \ne \varnothing$. Now

$$\begin{aligned} |\{|Sb| > \tfrac{1}{2}\lambda\}| &\le \Big| \Big\{ \Big| \sum_{L \in \mathscr{B}} \sum_{K \subseteq L} A_K b_L \Big| > 0 \Big\} \Big| + \Big| \Big\{ \Big| \sum_{L \in \mathscr{B}} \sum_{K \supsetneq L} A_K b_L \Big| > \tfrac{1}{2}\lambda \Big\} \Big| \\ &\le \sum_{L \in \mathscr{B}} |L| + \frac{2}{\lambda} \Big\| \sum_{L \in \mathscr{B}} \sum_{K \supsetneq L} A_K b_L \Big\|_1 \\ &\le \frac{1}{\lambda}\|f\|_1 + \frac{2}{\lambda} \sum_{L \in \mathscr{B}} \sum_{K \supsetneq L} \|A_K b_L\|_1, \end{aligned}$$

where we used the elementary properties of the Calderón–Zygmund decomposition to estimate the first term.

For the remaining double sum, we still need some observations. Recall that

$$A_K b_L = \sum_{\substack{I, J \subseteq K \\ \ell(I) = 2^{-i}\ell(K) \\ \ell(J) = 2^{-j}\ell(K)}} a_{IJK} h_I \langle h_J, b_L \rangle.$$

Now, if $\ell(K) > 2^\kappa \ell(L) \ge 2^j \ell(L)$, then $\ell(J) > \ell(L)$, and hence h_J is constant on L. But the integral of b_L vanishes, hence $\langle h_J, b_L \rangle = 0$ for all relevant J, and thus $A_K b_L = 0$ whenever $\ell(K) > 2^\kappa \ell(L)$.

Thus, in the inner sum, the only possible nonzero terms are $A_K b_L$ for $K = L^{(m)}$ for $m = 1, \ldots, \kappa$. By the separation of scales, at most one of these terms is nonzero, and we write $\tilde{L}$ for the corresponding unique K. So in fact

$$\frac{2}{\lambda} \sum_{L \in \mathscr{B}} \sum_{K \supsetneq L} \|A_K b_L\|_1 = \frac{2}{\lambda} \sum_{L \in \mathscr{B}} \|A_{\tilde{L}} b_L\|_1 \le \frac{2}{\lambda} \sum_{L \in \mathscr{B}} \|b_L\|_1 \le \frac{2}{\lambda} \cdot 2\|f\|_1 = \frac{4}{\lambda}\|f\|_1$$

by using the normalized boundedness of the averaging operators $A_{\tilde{L}}$ on $L^1(\mathbb{R}^d)$, and an elementary estimate for the bad part of the Calderón–Zygmund decomposition.

Altogether, we obtain the claim with $C = 4 \cdot 2^d + 5$. □

For the special subshifts $S_{\mathscr{K}_b^a(P)}$, we can improve the weak-type $(1,1)$ estimate to an exponential decay:

5.2. **Proposition.** *Let $S_{\mathscr{K}_b^a(P)}$ be the subshift of S as constructed earlier. Then the following estimate holds when ν is either the Lebesgue measure or w:*

$$\nu\Big(\Big\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)| > C2^{-b}\langle\sigma\rangle_P \cdot t\Big\}\Big) \lesssim C2^{-t}\nu(P), \qquad t \geq 0,$$

where C is a constant.

Proof. Let $\lambda := C2^{-b}\langle\sigma\rangle_P$, where C is a large constant, and $n \in \mathbb{Z}_+$. Let $x \in \mathbb{R}^d$ be a point where

$$|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)(x)| > n\lambda. \tag{5.3}$$

Then for all small enough $L \in \mathscr{K}_b^a(P)$ with $L \ni x$, there holds

$$\Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\supseteq L}} A_K(\sigma 1_Q)(x)\Big| > n\lambda.$$

Since $\displaystyle\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\supsetneq L}} A_K(\sigma 1_Q)$ is constant on L (thanks to separation of scales), and

$$\|A_L(\sigma 1_Q)\|_\infty \lesssim \frac{\sigma(L)}{|L|} \leq 2^{1-b}\frac{\sigma(P)}{|P|}, \tag{5.4}$$

it follows that

$$\Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\supsetneq L}} A_K(\sigma 1_Q)\Big| > (n - \tfrac{2}{3})\lambda \qquad \text{on } L. \tag{5.5}$$

Let $\mathscr{L} \subseteq \mathscr{K}_b^a(P)$ be the collection of maximal cubes with the above property. Thus all $L \in \mathscr{L}$ are disjoint, and all x with (5.3) belong to some L. By maximality of L, the minimal $L^* \in \mathscr{K}_b^a(S)$ with $L^* \supsetneq L$ satisfies

$$\Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\supsetneq L^*}} A_K(\sigma 1_Q)\Big| \leq (n - \tfrac{2}{3})\lambda \qquad \text{on } L^*.$$

By an estimate similar to (5.4), with L^* in place of L, it follows that

$$\Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\supsetneq L}} A_K(\sigma 1_Q)\Big| \leq (n - \tfrac{1}{3})\lambda \qquad \text{on } L.$$

Thus, if x satisfies (5.3) and $x \in L \in \mathscr{L}$, then necessarily

$$|S_{\{K\in\mathscr{K}_b^a(P);K\subseteq L\}}(\sigma 1_Q)(x)| = \Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\subseteq L}} A_K(\sigma 1_Q)(x)\Big| > \tfrac{1}{3}\lambda.$$

Using the weak-type L^1 estimate to the shift $S_{\{K\in\mathscr{K}_b^a(P);K\subseteq L\}}$ of type (i,j) with scales separated, noting that $A_K(\sigma 1_Q) = A_K(\sigma 1_L)$ for $K \subseteq L$, it follows that

$$\begin{aligned}\Big|\Big\{\Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\subseteq L}} A_K(\sigma 1_Q)(x)\Big| > \tfrac{1}{3}\lambda\Big\}\Big| &\leq \frac{C}{\lambda}\sigma(L) \\ &\leq \frac{C}{\lambda}2^{1-b}\frac{\sigma(S\cap Q)}{|S|}|L| \leq \tfrac{1}{3}|L|,\end{aligned}$$

provided that the constant in the definition of λ was chosen large enough. Recalling (5.5), there holds

$$\Big|\sum_{K\in\mathscr{K}_b^a(P)} A_K(\sigma 1_Q)\Big| \geq \Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\supsetneq L}} A_K(\sigma 1_Q)\Big| - \Big|\sum_{\substack{K\in\mathscr{K}_b^a(P)\\ K\subseteq L}} A_K(\sigma 1_Q)\Big|$$
$$> (n-\tfrac{2}{3})\lambda - \tfrac{1}{3}\lambda = (n-1)\lambda \quad \text{on } \tilde{L}\subset L \text{ with } |\tilde{L}|\geq \tfrac{2}{3}|L|.$$

Thus

$$\begin{aligned}
|\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)| > n\lambda\}| &\leq \sum_{L\in\mathscr{L}} |L\cap\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n\lambda\}| \\
&\leq \sum_{L\in\mathscr{L}} |\{|S_{\{K\in\mathscr{K}_b^a(P):K\subseteq L\}}(\sigma 1_Q)| > \tfrac{1}{3}\lambda\}| \\
&\leq \sum_{L\in\mathscr{L}} \tfrac{1}{3}|L| \leq \sum_{L\in\mathscr{L}} \tfrac{1}{3}\cdot\tfrac{3}{2}|\tilde{L}| \\
&\leq \tfrac{1}{2}\sum_{L\in\mathscr{L}} |L\cap\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>(n-1)\lambda\}| \\
&\leq \tfrac{1}{2}|\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>(n-1)\lambda\}|.
\end{aligned}$$

By induction it follows that

$$\begin{aligned}
|\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)| > n\lambda\}| &\leq 2^{-n}|\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>0\}| \\
&\leq 2^{-n}\sum_{M\in\mathscr{M}}|M| \leq 2^{-n}|P|,
\end{aligned}$$

where $\mathscr{M}$ is the collection of maximal cubes in $\mathscr{K}_b^a(S)$.

Recalling that we defined $\lambda := C2^{-b}\langle\sigma\rangle_P$ in the beginning of the proof, the previous display gives precisely the claim of the Proposition in the case that ν is the Lebesgue measure. We still need to consider the case that $\nu = w$. To this end, selected intermediate steps of the above computation, as well as the definition of $\mathscr{K}_b^a(P)$, will be exploited. Recall that $K\in\mathscr{K}^a$ means that $2^{a-1} < \langle w\rangle_K\langle\sigma\rangle_K \leq 2^a$, while $K\in\mathscr{K}_b^a(P)$ means that in addition $2^{-b} < \langle\sigma\rangle_K/\langle\sigma\rangle_P \leq 2^{1-b}$. Put together, this says that

$$2^{a+b-2}\langle\sigma\rangle_P < \frac{w(K)}{|K|} < 2^{a+b}\langle\sigma\rangle_P \qquad \forall K\in\mathscr{K}_b^a(P).$$

Hence, using the collections $\mathscr{L},\mathscr{M}\subseteq\mathscr{K}_b^a(P)$ as above,

$$\begin{aligned}
w(\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n\lambda\}) &\leq \sum_{L\in\mathscr{L}} w(L) \leq \sum_{L\in\mathscr{L}} 2^{a+b}\langle\sigma\rangle_P|L| \\
&\leq 2^{a+b}\langle\sigma\rangle_P|\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>(n-1)\lambda\}| \\
&\leq 2^{a+b}\langle\sigma\rangle_P\cdot 2^{-n}\sum_{M\in\mathscr{M}}|M| \\
&\leq 4\cdot 2^{-n}\sum_{M\in\mathscr{M}} w(M) \leq 4\cdot 2^{-n}w(S). \qquad \square
\end{aligned}$$

5.C. **Conclusion of the estimation of the testing conditions.** Recall that

$$\begin{aligned}&\Big\|\sum_{K\subseteq Q}A_K(\sigma 1_Q)\Big\|_{L^2(w)}\\&\leq\sum_{k=0}^{\kappa}\sum_a\sum_{b=0}^{\infty}2^{-b}\sum_{n=0}^{\infty}(1+n)\Big\|\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}\Big\|_{L^2(w)}\end{aligned}$$

and

$$\begin{aligned}&\Big\|\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}\Big\|_{L^2(w)}\\&\leq 2\Big(\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P^2 w(\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\})\Big)^{1/2}\\&\leq C\Big(\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P^2 2^{-n/C}w(P)\Big)^{1/2}\\&=C2^{-cn}\Big(\sum_{P\in\mathscr{P}^a}\frac{\sigma(P)w(P)}{|P|^2}\sigma(P)\Big)^{1/2}\\&\leq C2^{-cn}\Big(2^a\sum_{P\in\mathscr{P}^a}\sigma(P)\Big)^{1/2},\end{aligned}$$

recalling the freezing of the A_2 characteristic between 2^{a-1} and 2^a for cubes in $\mathscr{K}^a\supseteq\mathscr{P}^a$.

For the summation over the principal cubes, we observe that

$$\sum_{P\in\mathscr{P}^a}\sigma(P)=\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P|P|=\int_Q\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_P(x)\,\mathrm{d}x.$$

At any given x, if $P_0\subsetneq P_1\subsetneq\ldots\subseteq Q$ are the principal cubes containing it, we have

$$\sum_{P\in\mathscr{P}^a}\langle\sigma\rangle_P 1_P(x)=\sum_{m=0}^{\infty}\langle\sigma\rangle_{P_m}\leq\sum_{m=0}^{\infty}2^{-m}\langle\sigma\rangle_{P_0}=2\langle\sigma\rangle_{P_0}\leq 2M(\sigma 1_Q)(x),$$

where M is the dyadic maximal operator. Hence

$$\sum_{P\in\mathscr{P}^a}\sigma(P)\leq 2\int_Q M(\sigma 1_Q)\,\mathrm{d}x\leq 2[\sigma]_{A_\infty}\sigma(Q),$$

where we use the following notion of the A_∞ characteristic:

$$[\sigma]_{A_\infty}:=\sup_Q\frac{1}{\sigma(Q)}\int_Q M(\sigma 1_Q)\,\mathrm{d}x;$$

this was implicit already in the work of Fujii [11] and it was taken as an explicit definition by the author and C. Pérez [20].

Substituting back, we have

$$\begin{aligned}
&\Big\| \sum_{K\subseteq Q} A_K(\sigma 1_Q)\Big\|_{L^2(w)} \\
&\leq \sum_{k=0}^{\kappa}\sum_a\sum_{b=0}^{\infty} 2^{-b}\sum_{n=0}^{\infty}(1+n)\Big\| \sum_{P\in\mathscr{P}^a} \langle\sigma\rangle_P 1_{\{|S_{\mathscr{K}_b^a(P)}(\sigma 1_Q)|>n2^{-b}\langle\sigma\rangle_P\}}\Big\|_{L^2(w)} \\
&\leq \sum_{k=0}^{\kappa}\sum_a\sum_{b=0}^{\infty} 2^{-b}\sum_{n=0}^{\infty}(1+n)\cdot C2^{-cn}\Big(2^a\sum_{P\in\mathscr{P}^a}\sigma(P)\Big)^{1/2} \\
&\leq \sum_{k=0}^{\kappa}\sum_a\sum_{b=0}^{\infty} 2^{-b}\sum_{n=0}^{\infty}(1+n)\cdot C2^{-cn}\big(2^a[\sigma]_{A_\infty}\big)^{1/2} \\
&= C\cdot[\sigma]_{A_\infty}^{1/2}\sum_{k=0}^{\kappa}\Big(\sum_{a\leq\lceil\log_2[w,\sigma]_{A_2}\rceil} 2^{a/2}\Big)\Big(\sum_{b=0}^{\infty}2^{-b}\Big)\Big(\sum_{n=0}^{\infty}(1+n)\cdot 2^{-cn}\Big) \\
&\leq C\cdot[\sigma]_{A_\infty}^{1/2}\cdot(1+\kappa)\cdot[w,\sigma]_{A_2}^{1/2},
\end{aligned}$$

and thus the testing constant $\mathfrak{S}$ is estimated by

$$\mathfrak{S}\leq C\cdot(1+\kappa)\cdot[w,\sigma]_{A_2}^{1/2}\cdot[\sigma]_{A_\infty}^{1/2}.$$

By symmetry, exchanging the roles of w and σ, we also have the analogous result for $\mathfrak{S}^*$, and so we have completed the proof of the following:

5.6. **Theorem.** *Let $\sigma, w\in A_\infty$ be functions which satisfy the joint A_2 condition*

$$[w,\sigma]_{A_2} := \sup_Q \frac{w(Q)\sigma(Q)}{|Q|^2}<\infty.$$

Then the testing constants $\mathfrak{S}$ and $\mathfrak{S}^*$ associated with a dyadic shift S of type (i,j) satisfy the following bounds, where $\kappa:=\max\{i,j\}$:

$$\begin{aligned}
\mathfrak{S} &\leq C\cdot(1+\kappa)\cdot[w,\sigma]_{A_2}^{1/2}\cdot[\sigma]_{A_\infty}^{1/2},\\
\mathfrak{S}^* &\leq C\cdot(1+\kappa)\cdot[w,\sigma]_{A_2}^{1/2}\cdot[w]_{A_\infty}^{1/2}.
\end{aligned}$$

6. Conclusions

In this section we simply collect the fruits of the hard work done above. A combination of Theorems 4.2 and 5.6 gives the following two-weight inequality, whose qualitative version was pointed out by Lacey, Petermichl and Reguera [24]. In the precise form as stated, this result and its consequences below were obtained by Pérez and myself [20], although originally formulated only in the case that $\sigma = w^{-1}$ is the dual weight.

6.1. **Theorem.** *Let $\sigma, w\in A_\infty$ be functions which satisfy the joint A_2 condition*

$$[w,\sigma]_{A_2} := \sup_Q \frac{w(Q)\sigma(Q)}{|Q|^2}<\infty.$$

Then a dyadic shift S of type (i,j) satisfies $S(\sigma\cdot) : L^2(\sigma)\to L^2(w)$, and more precisely

$$\|S(\sigma\cdot)\|_{L^2(\sigma)\to L^2(w)}\lesssim (1+\kappa)^2[w,\sigma]_{A_2}^{1/2}\big([w]_{A_\infty}^{1/2}+[\sigma]_{A_\infty}^{1/2}\big),$$

where $\kappa=\max\{i,j\}$.

The quantitative bound as stated, including the polynomial dependence on κ, allows to sum up these estimates in the Dyadic Representation Theorem to deduce:

6.2. **Theorem.** *Let $\sigma, w \in A_\infty$ be functions which satisfy the joint A_2 condition. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0,1)$, satisfies*

$$\|T(\sigma\cdot)\|_{L^2(\sigma)\to L^2(w)} \lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w,\sigma]_{A_2}^{1/2}\big([w]_{A_\infty}^{1/2} + [\sigma]_{A_\infty}^{1/2}\big).$$

Recalling the dual weight trick and specializing to the one-weight situation with $\sigma = w^{-1}$, this in turn gives:

6.3. **Theorem.** *Let $w \in A_2$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0,1)$, satisfies*

$$\begin{aligned}\|T\|_{L^2(w)\to L^2(w)} &\lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w]_{A_2}^{1/2}\big([w]_{A_\infty}^{1/2} + [w^{-1}]_{A_\infty}^{1/2}\big)\\ &\lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w]_{A_2}.\end{aligned}$$

The second displayed line is the original A_2 theorem [15], and it follows from the first line by $[w]_{A_\infty} \lesssim [w]_{A_2}$ and $[w^{-1}]_{A_\infty} \lesssim [w^{-1}]_{A_2} = [w]_{A_2}$ (see Lemma 6.4 below). Its strengthening on the first line was first observed in my joint work with C. Pérez [20]. Note that, compared to the introductory statement in Theorem 1.1, the dependence on the operator T has been made more explicit. (The implied constants in the notation "$\lesssim$" only depend on the dimension and the Hölder exponent α.) This dependence on $\|T\|_{L^2\to L^2}$ and $\|K\|_{CZ_\alpha}$ is implicit in the original proof.

For completeness, we include the proof (in the stated form essentially from [24], but see also [20] for more general comparison of A_∞ and A_p constants) that

6.4. **Lemma.** *For all weights $w \in A_2$, we have*

$$[w]_{A_\infty} := \sup_Q \frac{1}{w(Q)} \int_Q M(1_Q w)\,\mathrm{d}x \le 8[w]_{A_2}.$$

Proof. Let $\mathcal{P}$ be the principal cubes of Muckenhoupt and Wheeden [31] given by $\mathcal{P} = \bigcup_{p=0}^\infty \mathcal{P}_p$, where $\mathcal{P}_0 := \{Q\}$ and $\mathcal{P}_{p+1}$ consists of the maximal $P' \subset P \in \mathcal{P}_p$ with $w(P')/|P'| > 2w(P)/|P|$. Then

$$M(1_Q w)(x) = \sup_{R:x\in R\subseteq Q} \frac{w(R)}{|R|} \le 2 \sup_{P\in\mathcal{P}:x\in P} \frac{w(P)}{|P|} \le 2\sum_{P\in\mathcal{P}} \frac{w(P)}{|P|} 1_P(x),$$

and hence

$$\int_Q M(1_Q w)\,\mathrm{d}x \le 2\sum_{P\in\mathcal{P}} w(P).$$

Consider the pairwise disjoint sets $E(P) := P \setminus \bigcup_{P'\in\mathcal{P}:P'\subsetneq P} P'$. Since

$$\sum_{\substack{P'\subsetneq P\\ P' \text{ maximal}}} |P'| \le \sum_{\substack{P'\subsetneq P\\ P' \text{ maximal}}} \frac{w(P')|P|}{2w(P)} \le \frac{w(P)|P|}{2w(P)} = \frac{|P|}{2},$$

it follows that $|E(P)| \geq \frac{1}{2}|P|$. We derive a similar condition for the weighted measure from the A_2 condition. Indeed,

$$\begin{aligned}
|E(P)| &= \int_{E(P)} w^{1/2} w^{-1/2}\,\mathrm{d}x \leq \Big(\int_{E(P)} w\,\mathrm{d}x\Big)^{1/2}\Big(\int_P w^{-1}\,\mathrm{d}x\Big)^{1/2} \\
&= w(E(P))^{1/2}\Big(\fint_P w^{-1}\,\mathrm{d}x\Big)^{1/2}|P|^{1/2} \\
&\leq w(E(P))^{1/2}[w]_{A_2}^{1/2}\Big(\fint_P w\,\mathrm{d}x\Big)^{-1/2}|P|^{1/2} = \Big([w]_{A_2}\frac{w(E(P))}{w(P)}\Big)^{1/2}|P|.
\end{aligned}$$

Using $|P| \leq 2|E(P)|$ and squaring, this shows that

$$w(P) \leq 4[w]_{A_2} w(E(P)).$$

After this, it is immediate to compute that

$$\sum_{P\in\mathcal{P}} w(P) \leq 4[w]_{A_2}\sum_{P\in\mathcal{P}} w(E(P)) \leq 4[w]_{A_2} w(Q),$$

since the sets $E(P)$ are pairwise disjoint and contained in Q. □

7. Further results and remarks

This final section briefly collects, without proofs, some further related developments, and poses some open problems.

The A_2 theorem implies a corresponding A_p theorem for all $p \in (1,\infty)$. This follows from a version of the celebrated extrapolation theorem, one of the most useful tools in the theory of A_p weights. The extrapolation theorem was first found by J. L. Rubio de Francia [42], and shortly after (so soon that it was published earlier) another proof was given by J. García-Cuerva [12]. For the present purposes, we need a quantitative form of the extrapolation theorem, which is due to Dragičević, Grafakos, Pereyra, and Petermichl [8], and reads as follows. Although relatively recent, it was known well before the proof of the full A_2 theorem.

7.1. **Theorem.** *If an operator T satisfies*

$$\|T\|_{L^2(w)\to L^2(w)} \leq C_T[w]_{A_2}^{\tau}$$

for all $w \in A_2$, then it satisfies

$$\|T\|_{L^p(w)\to L^p(w)} \leq c_p C_T[w]_{A_p}^{\tau\max\{1,1/(p-1)\}}$$

for all $p \in (1,\infty)$ and $w \in A_p$.

7.2. **Corollary.** *Let $p \in (1,\infty)$ and $w \in A_p$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0,1)$, satisfies*

$$\|T\|_{L^p(w)\to L^p(w)} \lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w]_{A_p}^{\max\{1,1/(p-1)\}}.$$

It is also possible to apply a version of the extrapolation argument to the mixed A_2/A_∞ bounds [20], but this did not give the optimal results for $p \neq 2$. However, by setting up a different argument directly in $L^p(w)$, the following bounds were obtained in my collaboration with M. Lacey [17]:

7.3. **Theorem.** *Let $p \in (1,\infty)$ and $w \in A_p$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0,1)$, satisfies*

$$\|T\|_{L^p(w)\to L^p(w)} \lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w]_{A_p}^{1/p}\big([w]_{A_\infty}^{1/p'} + [w^{1-p'}]_{A_\infty}^{1/p}\big).$$

For weak-type bounds, which were investigated by Lacey, Martikainen, Orponen, Reguera, Sawyer, Uriarte-Tuero, and myself [18], we need only 'half' of the strong-type upper bound:

7.4. **Theorem.** *Let $p \in (1,\infty)$ and $w \in A_p$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0,1)$, satisfies*

$$\begin{aligned}\|T\|_{L^p(w)\to L^{p,\infty}(w)} &\lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w]_{A_p}^{1/p}[w]_{A_\infty}^{1/p'} \\ &\lesssim (\|T\|_{L^2\to L^2} + \|K\|_{CZ_\alpha})[w]_{A_p}.\end{aligned}$$

All these results remain valid for the non-linear operators given by the *maximal truncations*

$$T_\natural f(x) := \sup_{\varepsilon>0}|T_\varepsilon f(x)|, \qquad T_\varepsilon f(x) := \int_{|x-y|>\varepsilon} K(x,y)f(y)\,\mathrm{d}y,$$

which have been addressed in [17, 18]. In [18] it was also shown that the sharp weighted bounds for dyadic shifts can be made linear (instead of quadratic) in κ, a result recovered by a different (Bellman function) method by Treil [43]. Earlier polynomial-in-κ Bellman function estimates for the shifts were due to Nazarov and Volberg [35]. An extension of the A_2 theorem to abstract metric spaces with a doubling measure (spaces of homogeneous type) is due to Nazarov, Reznikov, and Volberg [32].

A higher degree of non-linearity is obtained by replacing the supremum over $\epsilon > 0$ defining the maximal truncation by one of the *variation norms*

$$\|\{v_\epsilon\}_{\epsilon>0}\|_{V^q} := \sup_{\{\epsilon_i\}_{i\in\mathbb{Z}}}\Big(\sum_i |v_{\epsilon_i} - v_{\epsilon_{i+1}}|^q\Big)^{1/q},$$

where the supremum is over all monotone sequences $\{\epsilon_i\}_{i\in\mathbb{Z}} \subset (0,\infty)$. Sharp weighted bounds for the q-variation ($q \in (2,\infty)$) of Calderón–Zygmund operators were first proved by Hytönen–Lacey–Pérez [19], although replacing the sharp truncation $T_\epsilon f(x)$ by a smooth truncation

$$T_\epsilon^\phi f(x) := \int \phi\Big(\frac{|x-y|}{\epsilon}\Big)K(x,y)f(y)\,\mathrm{d}y,$$

where ϕ is smooth and $0 \le \phi \le 1_{(1,\infty)}$. Sharp weighted bounds for the q-variation of the sharp truncations with $\phi = 1_{(1,\infty)}$ were recently obtained by de França Silva and Zorin-Kranich [6].

The approach to the q-variation in [19] was through a non-probabilistic counterpart of the Dyadic Representation, a Dyadic Domination, which was independently discovered by Lerner [26, 27]. Another advantage of this method was its ability to handle Calderón–Zygmund kernels with weaker moduli of continuity ψ than those treated by the present approach; namely any moduli ψ subject to the log-bumped Dini condition $\int_0^1 \psi(t)(1+\log\frac{1}{t})\frac{\mathrm{d}t}{t} < \infty$.

In its original form, the Dyadic Domination theorem gave a domination in norm, which improved to pointwise domination by Conde-Alonso and Rey [2] and, independently, by Lerner and Nazarov [29]. All these approaches required the same log-Dini condition, and the necessity of the logarithmic correction to the Dini-condition remained open for some time, until it was finally eliminated by Lacey [23] by yet another approach. The following quantitative form of Lacey's theorem was obtained by L. Roncal, O. Tapiola and the author [22], and with a simpler proof by Lerner [28]:

7.5. **Theorem.** *Let $w \in A_2$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has modulus of continuity ψ , satisfies*

$$\|T\|_{L^2(w)\to L^2(w)} \lesssim \Big(\|T\|_{L^2\to L^2} + \|K\|_{CZ_0} + \|K\|_{CZ_\psi}\int_0^1 \psi(t)\frac{\mathrm{d}t}{t}\Big)[w]_{A_2}.$$

Asking for even less regularity, one may wonder about the sharp weighted bound for the class of rough homogeneous singular integral operators

$$Tf(x) = \text{p.v.}\int_{\mathbb{R}^d}\frac{\Omega(y)}{|y|^d}f(x-y)\,\mathrm{d}y,$$

where

$$\Omega(y) = \Omega(\frac{y}{|y|}), \qquad \Omega \in L^\infty(\mathbb{S}^{d-1}), \qquad \int_{\mathbb{S}^{d-1}}\Omega(\sigma)\,\mathrm{d}\sigma = 0.$$

Their qualitative boundedness $T : L^2(w) \to L^2(w)$ is known for $w \in A_2$ (see Watson [45]). Roncal, Tapiola and the author [22] showed that $\|T\|_{L^2(w)\to L^2(w)} \lesssim \|\Omega\|_\infty[w]_{A_2}^2$, but it is not known whether this quadratic dependence on $[w]_{A_2}$ is sharp.

7.A. **The Beurling operator and its powers.** One of the key original motivations to study the A_2 theorem was a conjecture of Astala–Iwaniec–Saksman [1] concerning the special case where T is the Beurling operator

$$Bf(z) := -\frac{1}{\pi}\,\text{p. v.}\int_{\mathbb{C}}\frac{1}{\zeta^2}f(z-\zeta)\,\mathrm{d}A(\zeta),$$

and A is the area measure (two-dimensional Lebesgue measure) on $\mathbb{C} \simeq \mathbb{R}^2$. This was the first Calderón–Zygmund operator for which the A_2 theorem was proven; it was achieved by Petermichl and Volberg [40], confirming the mentioned conjecture of Astala, Iwaniec, and Saksman [1]. Another proof of the A_2 theorem for this specific operator is due to Dragičević and Volberg [10].

The powers B^n of B have also been studied, and then it is of interest to understand the growth of the norms as a function of n. Shortly before the proof of the full A_2 theorem, by methods specific to the Beurling operator, O. Dragičević [7] was able to prove the cubic growth

$$\|B^n\|_{L^2(w)\to L^2(w)} \lesssim |n|^3[w]_{A_2}, \qquad n \in \mathbb{Z}\setminus\{0\}.$$

Now, let us see what the general A_2 theorem gives for these specific powers. It is known (see e.g. [9]) that B^n is the convolution operator with the kernel

$$K_n(z) = (-1)^n\frac{|n|}{\pi}\Big(\frac{\bar{z}}{z}\Big)^n|z|^{-2},$$

and it is elementary to check that this satisfies $\|K_n\|_{CZ_\alpha} \lesssim |n|^{1+\alpha}$ for any $\alpha \in (0,1)$. Moreover, since B is an isometry on $L^2(\mathbb{C})$, we have $\|B^n\|_{L^2\to L^2} = 1$. From Theorem 6.3 we deduce:

7.6. **Corollary.** *The powers B^n of the Beurling operator satisfy*

$$\|B^n\|_{L^2(w)\to L^2(w)} \lesssim |n|^{1+\alpha}[w]_{A_2}, \qquad \alpha > 0,$$

where the implied constant depends on α.

A sharper estimate still is provided by Theorem 7.5, as observed in [22]:

7.7. **Corollary.** *The powers B^n of the Beurling operator satisfy*

$$\|B^n\|_{L^2(w)\to L^2(w)} \lesssim |n|(1+\log|n|)[w]_{A_2}.$$

For this it suffices to check that, defining the modulus of continuity

$$\psi_n(t) := \min\{|n|t, 1\},$$

we have $\|K_n\|_{CZ_{\psi_n}} \lesssim |n|$ and hence

$$\|K_n\|_{CZ_{\psi_n}} \int_0^1 \psi_n(t)\frac{\mathrm{d}t}{t} \lesssim |n|(1+\log|n|).$$

However, a better bound would follow if we had the A_2 theorem for the rough singular integrals in the form

$$\|T\|_{L^2(w)\to L^2(w)} \lesssim \|\Omega\|_\infty [w]_{A_2},$$

for this would lead to the linear estimate $\|B^n\|_{L^2(w)\to L^2(w)} \lesssim |n|[w]_{A_2}$, simply by viewing the kernels K_n (although smooth), as rough kernels of homogeneous singular integrals.

Let us notice that no bound better than this is possible, at least on the scale of power-type dependence on $|n|$:

7.8. **Proposition.** *No bound of the form $\|B^n\|_{L^2(w)\to L^2(w)} \lesssim |n|^{1-\epsilon}[w]_{A_2}^\tau$ can be valid for any $\epsilon, \tau > 0$.*

Proof. Suppose for contradiction that such a bound holds for some fixed $\epsilon, \tau > 0$ and all $n \in \mathbb{Z}\setminus\{0\}$. By Theorem 7.1, we deduce that

$$\|B^n\|_{L^p(w)\to L^p(w)} \lesssim_p |n|^{1-\epsilon}[w]_{A_p}^{\tau\max\{1,1/(p-1)\}},$$

and hence in particular we have the unweighted bound

$$\|B^n\|_{L^p\to L^p} \lesssim_p |n|^{1-\epsilon}, \qquad 1 < p < \infty.$$

However, it has been shown by Dragičević, Petermichl and Volberg that the correct dependence here is

$$\|B^n\|_{L^p\to L^p} \eqsim_p |n|^{|1-2/p|}, \qquad 1 < p < \infty.$$

The previous two displays are clearly in contradiction for p close to either 1 or ∞, and we are done. $\square$

The quest for the A_2 theorem began from the investigations of the Beurling transform, but clearly even this case is not yet fully understood.

Acknowledgements

The author was supported by the European Union via the ERC Starting Grant "Analytic–probabilistic methods for borderline singular integrals", and by the Academy of Finland via grants 130166 and 133264 and the Centre of Excellence in Analysis and Dynamics Research. The core of this exposition was first presented as a series of lectures during the Doc-Course "Harmonic analysis, metric spaces and applications to P.D.E." at Universidad de Sevilla, Spain, in June-July 2011. I would like to thank the organizers of the summer school, and especially Carlos Pérez, for this opportunity. I would also like to thank Oliver Dragičević for discussions that led me to the observations about the powers of the Beurling operator in Section 7.A. The anonymous referee is warmly thanked for his/her friendly suggestions that eliminated several serious omissions in my original list of citations.

References

[1] K. Astala, T. Iwaniec, and E. Saksman. Beltrami operators in the plane. *Duke Math. J.*, 107(1):27–56, 2001.
[2] J. M. Conde-Alonso and G. Rey. A pointwise estimate for positive dyadic shifts and some applications. *Math. Ann.*, 365(3-4):1111–1135, 2016.
[3] D. Cruz-Uribe, J. M. Martell, and C. Pérez. Sharp weighted estimates for classical operators. *Adv. Math.*, 229(1):408–441, 2012.
[4] L. Dalenc and Y. Ou. Upper bound for multi-parameter iterated commutators. *Publ. Mat.*, 60(1):191–220, 2016.
[5] G. David and J.-L. Journé. A boundedness criterion for generalized Calderón-Zygmund operators. *Ann. of Math. (2)*, 120(2):371–397, 1984.
[6] F. C. de França Silva and P. Zorin-Kranich. Sparse domination of sharp variational truncations. *ArXiv e-prints*, Apr. 2016.
[7] O. Dragičević. Weighted estimates for powers and the Ahlfors-Beurling operator. *Proc. Amer. Math. Soc.*, 139(6):2113–2120, 2011.
[8] O. Dragičević, L. Grafakos, M. C. Pereyra, and S. Petermichl. Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces. *Publ. Mat.*, 49(1):73–91, 2005.
[9] O. Dragičević, S. Petermichl, and A. Volberg. A rotation method which gives linear L^p estimates for powers of the Ahlfors-Beurling operator. *J. Math. Pures Appl. (9)*, 86(6):492–509, 2006.
[10] O. Dragičević and A. Volberg. Sharp estimate of the Ahlfors-Beurling operator via averaging martingale transforms. *Michigan Math. J.*, 51(2):415–435, 2003.
[11] N. Fujii. Weighted bounded mean oscillation and singular integrals. *Math. Japon.*, 22(5):529–534, 1977/78.
[12] J. García-Cuerva. An extrapolation theorem in the theory of A_p weights. *Proc. Amer. Math. Soc.*, 87(3):422–426, 1983.
[13] I. Holmes, M. T. Lacey, and B. D. Wick. Commutators in the Two-Weight Setting. *ArXiv e-prints*, June 2015.
[14] I. Holmes and B. D. Wick. Two Weight Inequalities for Iterated Commutators with Calder\'on-Zygmund Operators. *ArXiv e-prints*, Sept. 2015.
[15] T. Hytönen. The sharp weighted bound for general Calderón-Zygmund operators. *Ann. of Math. (2)*, 175(3):1473–1506, 2012.
[16] T. Hytönen. The vector-valued nonhomogeneous Tb theorem. *Int. Math. Res. Not. IMRN*, (2):451–511, 2014.
[17] T. Hytönen and M. T. Lacey. The A_p-A_∞ inequality for general Calderón-Zygmund operators. *Indiana Univ. Math. J.*, 61(6):2041–2092, 2012.
[18] T. Hytönen, M. T. Lacey, H. Martikainen, T. Orponen, M. C. Reguera, E. T. Sawyer, and I. Uriarte-Tuero. Weak and strong type estimates for maximal truncations of Calderón-Zygmund operators on A_p weighted spaces. *J. Anal. Math.*, 118(1):177–220, 2012.
[19] T. Hytönen, M. T. Lacey, and C. Pérez. Sharp weighted bounds for the q-variation of singular integrals. *Bull. Lond. Math. Soc.*, 45(3):529–540, 2013.

[20] T. Hytönen and C. Pérez. Sharp weighted bounds involving A_∞. *Anal. PDE*, 6(4):777–818, 2013.
[21] T. Hytönen, C. Pérez, S. Treil, and A. Volberg. Sharp weighted estimates for dyadic shifts and the A_2 conjecture. *J. Reine Angew. Math.*, 687:43–86, 2014.
[22] T. Hytönen, L. Roncal, and O. Tapiola. Quantitative weighted estimates for rough homogeneous singular integrals. *ArXiv e-prints*, Oct. 2015.
[23] M. T. Lacey. An elementary proof of the A_2 Bound. *ArXiv e-prints*, Jan. 2015.
[24] M. T. Lacey, S. Petermichl, and M. C. Reguera. Sharp A_2 inequality for Haar shift operators. *Math. Ann.*, 348(1):127–141, 2010.
[25] A. K. Lerner. A pointwise estimate for the local sharp maximal function with applications to singular integrals. *Bull. Lond. Math. Soc.*, 42(5):843–856, 2010.
[26] A. K. Lerner. On an estimate of Calderón-Zygmund operators by dyadic positive operators. *J. Anal. Math.*, 121:141–161, 2013.
[27] A. K. Lerner. A simple proof of the A_2 conjecture. *Int. Math. Res. Not. IMRN*, (14):3159–3170, 2013.
[28] A. K. Lerner. On pointwise estimates involving sparse operators. *New York J. Math.*, 22:341–349, 2016.
[29] A. K. Lerner and F. Nazarov. Intuitive dyadic calculus: the basics. *ArXiv e-prints*, Aug. 2015.
[30] H. Martikainen. Representation of bi-parameter singular integrals by dyadic operators. *Adv. Math.*, 229(3):1734–1761, 2012.
[31] B. Muckenhoupt and R. L. Wheeden. Some weighted weak-type inequalities for the Hardy-Littlewood maximal function and the Hilbert transform. *Indiana Univ. Math. J.*, 26(5):801–816, 1977.
[32] F. Nazarov, A. Reznikov, and A. Volberg. The proof of A_2 conjecture in a geometrically doubling metric space. *Indiana Univ. Math. J.*, 62(5):1503–1533, 2013.
[33] F. Nazarov, S. Treil, and A. Volberg. The Tb-theorem on non-homogeneous spaces. *Acta Math.*, 190(2):151–239, 2003.
[34] F. Nazarov, S. Treil, and A. Volberg. Two weight inequalities for individual Haar multipliers and other well localized operators. *Math. Res. Lett.*, 15(3):583–597, 2008.
[35] F. Nazarov and A. Volberg. A simple sharp weighted estimate of the dyadic shifts on metric spaces with geometric doubling. *Int. Math. Res. Not. IMRN*, (16):3771–3789, 2013.
[36] Y. Ou. A $T(b)$ theorem on product spaces. *Trans. Amer. Math. Soc.*, 367(9):6159–6197, 2015.
[37] C. Perez, S. Treil, and A. Volberg. On A_2 conjecture and corona decomposition of weights. *ArXiv e-prints*, June 2010.
[38] S. Petermichl. The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical A_p characteristic. *Amer. J. Math.*, 129(5):1355–1375, 2007.
[39] S. Petermichl. The sharp weighted bound for the Riesz transforms. *Proc. Amer. Math. Soc.*, 136(4):1237–1249, 2008.
[40] S. Petermichl and A. Volberg. Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular. *Duke Math. J.*, 112(2):281–305, 2002.
[41] S. Pott and A. Stoica. Linear bounds for Calderón-Zygmund operators with even kernel on UMD spaces. *J. Funct. Anal.*, 266(5):3303–3319, 2014.
[42] J. L. Rubio de Francia. Factorization theory and A_p weights. *Amer. J. Math.*, 106(3):533–547, 1984.
[43] S. Treil. Sharp A_2 estimates of Haar shifts via Bellman function. *ArXiv e-prints*, May 2011.
[44] A. Vagharshakyan. Recovering singular integrals from Haar shifts. *Proc. Amer. Math. Soc.*, 138(12):4303–4309, 2010.
[45] D. K. Watson. Weighted estimates for singular integrals via Fourier transform estimates. *Duke Math. J.*, 60(2):389–399, 1990.

Department of Mathematics and Statistics, P.O.B. 68 (Gustaf Hällströmin katu 2b), FI-00014 University of Helsinki, Finland
E-mail address: `tuomas.hytonen@helsinki.fi`