SHARP $A_2$ ESTIMATES OF HAAR SHIFTS VIA BELLMAN FUNCTION

SERGEI TREIL

Abstract. We use the Bellman function method to give an elementary proof of a sharp weighted estimate for the Haar shifts, which is linear in the $A_2$ norm of the weight and in the complexity of the shift. Together with the representation of a general Calderón–Zygmund operator as a weighted average (over all dyadic lattices) of Haar shifts, cf. [10, 11], it gives a significantly simpler proof of the so-called the $A_2$ conjecture.

The main estimate (Lemma 6.1) is a very general fact about concave functions, which can be very useful in other problems of martingale Harmonic Analysis. Concave functions of such type appear as the Bellman functions for bounds on the bilinear form of martingale multipliers, thus the main estimate allows for the transference of the results for simplest possible martingale multipliers to more general martingale transforms.

Note that (although this is not important for the $A_2$ conjecture for general Calderón–Zygmund operators) this elementary proof gives the best known (linear) growth in the complexity of the shift.

Contents

Notation 2
1. Introduction 2
2. Main objects. 4
3. Bellman function for sharp weighted estimates of the dyadic martingale multipliers 6
4. The main result: sharp weighted estimates of the Haar shifts 9
5. Preliminaries for the proof 9
6. The main estimate 11
7. Proof of the main result (Theorem 4.1) 17
8. Estimates of the paraproducts 18
9. Concluding remarks 21
References 22

2010 Mathematics Subject Classification. 42B20, 42B35, 26B25, 60G42, 60G46.
Key words and phrases. Calderón–Zygmund operators, $A_2$ weights, Haar shift, dyadic shift, Bellman function, non-homogeneous Harmonic Analysis, Harmonic Analysis on martingales.

Supported by the National Science Foundation under the grant DMS-0800876.
Notation

- $\mathcal{D}$: a dyadic lattice in $\mathbb{R}$ or $\mathbb{R}^d$;
- $\text{chld } I$: the collection of children of the interval (cube) $I$;
- $\text{chld}_k I$: the collection of children of the order $k$ of the interval (cube) $I$; the collection $\text{chld}_0(I)$ consist of the interval $I$;
- $|E|$: the $d$-dimensional Lebesgue measure of a set $E \subset \mathbb{R}^d$;
- $\langle f \rangle_I$, $\mathbb{E}_T f$: average of $f$ over $I$, $\langle f \rangle_I = |I|^{-1} \int_I f(x) dx$;
- $E_I$: averaging operator, $E_I f := \langle f \rangle_I 1_I$;
- $\Delta_I$: Martingale difference operator, $\Delta_I := -E_I + \sum_{J \in \text{chld}(I)} E_J$;
- $L^2(w)$: the weighted $L^2$ space, $\|f\|_{L^2(w)}^2 = \int_{\mathbb{R}^n} |f(x)|^2 w(x) dx$.

1. Introduction

The famous Hunt–Muckenhoupt–Wheeden theorem states that a Calderón–Zygmund operator $T$ is bounded in the weighted space $L^2(w) = L^2(\mathbb{R}^d, w)$ if and only if the weight $w$ satisfies the so-called Muckenhoupt condition

$$(A_2) \quad \sup_Q \left( |Q|^{-1} \int_Q w dx \right) \left( |Q|^{-1} \int_Q w^{-1} dx \right) =: [w]_{A_2} < \infty,$$

where the supremum is taken over all cubes $Q$ in $\mathbb{R}^d$. The quantity $[w]_{A_2}$ is called the Muckenhoupt (or $A_2$) norm of the weight $w$ (although it is clearly not a norm).

It has been an old problem to describe how the norm of a Calderón–Zygmund operator in the weighted space $L^2(w)$ depends on the Muckenhoupt norm $[w]_{A_2}$ of $w$. A conjecture was that for a fixed Calderón–Zygmund operator $T$ its norm in $L^2(w)$ is bounded by $C \cdot [w]_{A_2}$, where the constant $C$ depends on the operator $T$ (but not on the weight $w$). Simple counterexamples demonstrate that for the classical operators like Hilbert Transform or Riesz Transform, a better estimate than $C \cdot [w]_{A_2}$ is generally not possible.

For operators that are not “too singular” better than linear estimates are possible. For example, for the convolution with the Poisson kernel $P_\tau$ its norm in $L^2(w)$ is estimated by $C[w]_{A_2}^{1/2}$, i.e.

$$\|P_\tau * f\|_{L^2(w)} \leq C[w]_{A_2}^{1/2} \|f\|_{L^2(w)}.$$

This can be easily explained: the norm of the averaging operator $E_I$, $E_I f = \langle f \rangle_I 1_I$ is exactly $\langle w \rangle_I^{1/2} \langle w^{-1} \rangle_I^{1/2}$, and the convolution with the Poisson kernel can be estimated by an average of the averaging operators. Also, it is not hard to show that this bound is sharp, meaning that for each weight $w$ convolution with some $P_\tau$ has the norm at least $c[w]_{A_2}^{1/2}$.

But for the classical singular integral operators like Hilbert Transform or Riesz Transforms the linear in $[w]_{A_2}$ estimate is the best one can hope to get. The conjecture about a linear in $[w]_{A_2}$ estimate of the norm for such and more general operators has become known as the $A_2$ conjecture.
For the maximal function, the estimate $C \cdot [w]_{A_2}$ was proved by S. Buckley [5]; he also proved that this estimate is optimal for the maximal function. The first result for a “singular integral” operator was due to J. Wittwer [27], who proved the $A_2$ conjecture for the Haar multipliers.

Using this result and Bellman function technique S. Petermichl and A. Volberg [23] proved the $A_2$ conjecture for the Beurling–Ahlfors operator (convolution with $\pi^{-1} z^{-2}$ in $\mathbb{C}$). An alternative proof was given a little later by O. Dragičević and A. Volberg [7] via the representation of the Beurling–Ahlfors Transform as an average of Haar multipliers over all dyadic lattices.

The result for other Calderón–Zygmund operators, in particular for the Hilbert Transform (the convolution with $(\pi x)^{-1}$ on $\mathbb{R}$) remained open for some time.

The next breakthrough was made by S. Petermichl [20] who proved the $A_2$ conjecture for the Hilbert Transform. She used the representation of the Hilbert Transform as the average over all (translated and dilated dyadic lattices) of a simple dyadic transformation (the so called Haar shift), and proved the $A_2$ conjecture for this operator. Later [21] she used similar ideas to prove the $A_2$ conjecture for the Riesz Transforms.

Note that it took a lot of time and effort to go from J. Wittwer’s [27] estimate for the simplest dyadic operator to the S. Petermichl’s [20] estimate for the Haar shift, despite the fact that the Haar shift was just a little bit more complicated than the Haar multiplier.

Lately, there was a lot of activity, that eventually lead to the complete solution of the $A_2$ conjecture for all Calderón–Zygmund operators. Namely, M. Lacey, S. Petermichl and M. Reguera [13] proved the $A_2$ conjecture for all so-called Haar shifts, see Definitions 2.1, 2.2 below, generalizations of the operators used in [20]. This result immediately implies $A_2$ conjecture for all operators that can be represented as averages of Haar shifts of fixed complexity (see Definition 2.1 below), in particular it gave another proof for Riesz transforms in $\mathbb{R}^d$.

The technique used in [13] was a very clever application of the stopping moment reasonings, and was based on the result from [17] about two-weight estimates for the so-called well-localized operators. Unfortunately, the estimates of the norm in [13] grew exponentially in the complexity of the Haar shift, so $A_2$ conjecture for general Calderón–Zygmund operators remained open.

For general Calderón–Zygmund operators the conjecture was finally settled by T. Hytönen [10]. One of the crucial components in his proof was the representation of an arbitrary Calderón–Zygmund operator as a weighted average (over all translated dyadic lattices) of the Haar shifts (of all possible complexities), where the weights decrease exponentially in the complexity. This would imply the $A_2$ conjecture if one could prove the $A_2$ conjecture for the general Haar shifts with the estimates depending sub-exponentially (for example polynomially) on the complexity.

\[1\text{In fact, for general Calderón–Zygmund operators one also should add the averages of so-called paraproducts and their adjoint. But the paraproducts are the operators of fixed complexity, so any known linear in $[w]_{A_2}$ estimate for such operators would work. We say more about estimates for paraproducts later in Section 8.}\]
This program was realized later in [11], where the polynomial in complexity (and of course linear in \([w]_{A_2}\)) estimate of the norm was obtained for general Haar shifts; a simpler representation of a Calderón–Zygmund operator as a weighted average of the Haar shifts was also presented there.

The original proof in [10] used a result from a very technical paper [19], where the \(A_2\) conjecture was reduced to a Sawyer type testing condition and as a corollary to a weak type estimate.

We should also mention as an interesting fact that all the proofs mentioned above used the results and/or technique from two papers by Nazarov–Treil–Volberg [16, 18] about two weight estimates for martingale transforms.

In the present paper we present a simple proof of the \(A_2\) conjecture for arbitrary Haar shifts with linear in complexity estimates. We get this estimate directly from the J. Wittwer result [27] (\(A_2\) conjecture for the simplest Haar multipliers) using the Bellman function technique. Together with the representation of a Calderón–Zygmund operator as a weighted average of Haar shifts from [10] or [11] it gives the \(A_2\) conjecture for general Calderón–Zygmund operators.

The proof is really simple and elementary; it is significantly simpler than any previous proof for the Haar shifts. It is really a shame that this proof was not discovered earlier.

The essence of the proof is a simple result about convex functions. Functions of such type appear often when one uses Bellman function method in dyadic Harmonic analysis, so the main result can be used for the transference, when one extends result obtained for the simplest dyadic model to more general martingale ones.

Finally, we should also mention preprint [24] where the Bellman function method was used to prove the estimates for the Haar shifts of complexity 0 and 1 (in particular for the S. Petermichl’s Haar shifts [20]).

2. Main objects.

2.1. Dyadic lattices. The standard dyadic system in \(\mathbb{R}^d\) is

\[
\mathcal{D}^0 := \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k^0, \quad \mathcal{D}_k^0 := \{2^k([0,1]^d + m) : m \in \mathbb{Z}^d\}.
\]

For \(I \in \mathcal{D}_k^0\) and a binary sequence \(\omega = (\omega_j)_{j=-\infty}^{\infty} \in \{(0,1)^d\}^\mathbb{Z}\), let

\[
I \dot{+} \omega := I + \sum_{j < k} \omega_j 2^j.
\]

Following Nazarov, Treil and Volberg [17, Section 9.1], consider general dyadic systems of the form

\[
\mathcal{D} = \mathcal{D}^\omega := \{I \dot{+} \omega : I \in \mathcal{D}^0\} = \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k^\omega.
\]

Given a cube \(I = x + [0, \ell]^d\), let

\[
\text{chld}(I) := \{x + \eta \ell/2 + [0, \ell/2]^d : \eta \in \{0,1\}^d\}
\]
denote the collection of dyadic children of $I$. Thus $\mathcal{D}_{k-1} = \bigcup \{ \text{child}(I) : I \in \mathcal{D}_k \}$. Note that, in line with [17, 11] but contrary to [10], we use the “geometric” indexing of cubes, where larger $k$ refers to larger cubes, rather than the “probabilistic” indexing, where larger $k$ would refer to finer sigma-algebras.

2.2. Martingale difference decompositions and Haar functions. For a cube $I$ in $\mathbb{R}^d$ let

$$E_I f := \left( \int_I f \, dx \right) 1_I := \left( |I|^{-1} \int_I f \, dx \right) 1_I, \quad \Delta_I := -E_I + \sum_{J \in \text{child}(I)} E_J.$$

It is well known that for an arbitrary dyadic lattice $\mathcal{D}$ every function $f \in L^2(\mathbb{R}^d)$ admits the orthogonal decomposition

$$f = \sum_{I \in \mathcal{D}} \Delta_I f.$$

Given a cube $Q$ in $\mathbb{R}^d$, any function in the martingale difference space $\Delta_Q L^2$ is called a Haar function (corresponding to $Q$) and is usually denoted by $h_Q$. Note, that here $h_Q$ denotes a *generic* Haar function, not any particular one.

In other words, a Haar function $h_Q$ is supported on $Q$, constant on the children of $Q$ and orthogonal to constants.

2.3. Dyadic shifts.

**Definition 2.1.** Let $m, n \in \mathbb{N}$. According to [11] an elementary dyadic shift with parameters $m, n$ is an operator given by

$$Sf := \sum_{Q \in \mathcal{D}} \sum_{Q', Q'' \in \mathcal{D}} |Q|^{-1} (f, h_{Q''}^{Q'}) h_{Q'}^{Q''},$$

where $h_{Q''}^{Q'}$ and $h_{Q'}^{Q''}$ are (non-weighted) Haar functions for the cubes $Q'$ and $Q''$ respectively, subject to normalization

$$(2.1) \quad \|h_{Q'}^{Q''}\|_\infty \cdot \|h_{Q''}^{Q'}\|_\infty \leq 1.$$

Notice that this implies, in particular, that

$$(2.2) \quad Sf(x) = \sum_{Q \in \mathcal{D}} |Q|^{-1} \int_Q a_Q(x, y) f(y) \, dy, \quad \text{supp } a_Q \subset Q \times Q, \quad \|a_Q\|_\infty \leq 1,$$

where

$$(2.3) \quad a_Q(x, y) = \sum_{Q', Q'' \in \mathcal{D}, Q', Q'' \subset Q, \ell(Q') = 2^{-m} \ell(Q), \ell(Q'') = 2^{-n} \ell(Q)} h_{Q'}^{Q''}(x) h_{Q''}^{Q'}(y).$$

We will call the number $\max(m, n) + 1$ the *complexity* of the dyadic shift. Note that in [10, 11] the complexity was defined as $\max(m, n)$; we use $\max(m, n) + 1$, because it will be more convenient for our purposes.
We will use a little more general definition of a Haar shift.

**Definition 2.2.** A Haar shift $S$ of complexity $n$ is given by

$$Sf = \sum_{Q \in \mathcal{D}} S_Q \Delta^n_Q f,$$

where the operators $S_Q$ act on $\Delta^n_Q L^2$ and can be represented as integral operators with kernels $a_Q$, $\|a_Q\|_\infty \leq |Q|^{-1}$. The latter means that for all $f, g \in \Delta^n_Q L^2$

$$\langle S_Q f, g \rangle = \int_Q a_Q(x, y) f(y) g(x) dxdy.$$

We can always think that our dyadic shifts $S$ are finite dyadic shifts meaning that only finitely many $Q$’s are involved in its definition above. All estimates will be independent of this finite number.

### 3. Bellman function for sharp weighted estimates of the dyadic martingale multipliers

The simplest example of a Haar shift is the so-called dyadic martingale multiplier. Namely, let $\mathcal{D}$ be the standard dyadic lattice in $\mathbb{R}$. For an interval $I \in \mathcal{D}$ let $h_I$ be the standard $L^2$-normalized Haar function, $h_I := |I|^{-1/2}(1_{I_1} - 1_{I_2})$, where $I_1$ and $I_2$ are the left and the right halves of $I$ respectively. Given a numerical sequence $\sigma = \{\sigma_I\}_{I \in \mathcal{D}}$, $|\sigma_I| \leq 1$, define the operator $T_\sigma$ by

$$T_\sigma f = \sum_{I \in \mathcal{D}} \sigma_I \langle f, h_I \rangle h_I.$$

The family of operators $T_\sigma$ can be considered to be the simplest dyadic analog of singular integral operators, so the sharp weighted estimates for such operators were the natural thing to try before attacking the case of general Calderón–Zygmund operators.

It was shown by J. Wittwer [27] that if a weight $w$ satisfies the dyadic Muckenhoupt condition

$$\sup_{I \in \mathcal{D}} \langle w \rangle_I (w^{-1})_I =: [w]_{A_2}^d < \infty$$

the operators $T_\sigma$ are (uniformly in $\sigma$, $|\sigma_I| \leq 1$) bounded in $L^2(w)$ by $C_1 [w]_{A_2}^d$, where $C_1$ is an absolute constant.

Denoting by $I_1$ and $I_2$ he children of $I$ one can rewrite the estimate of the norm as

$$\sum_{I \in \mathcal{D}} \langle f \rangle_{I_1} - \langle f \rangle_{I_2} \cdot |g|_{I_1} - |g|_{I_2} \cdot |I| \leq C [w]_{A_2}^d \|f\|_{L^2(w)} \|g\|_{L^2(w^{-1})},$$

for all $f \in L^2(w)$, $g \in L^2(w^{-1})$; here $C = C_1/4.$
3.1. Bellman function for the martingale multipliers and its properties. Following the standard Bellman function technique, cf. [14, 15], let us define the Bellman function for the problem. Let $A > 1$, and let $I_1$ and $I_2$ denote the children of an interval $I$. Fix a dyadic interval $I_0$, for example $I_0 = [0, 1]$ and for real numbers $f, g, F, G, u, v$ satisfying

\[(3.2)\]

$u, v > 0, \quad 1 \leq uv \leq A, \quad f^2 \leq Fv, \quad g^2 \leq Gu$

define the function $B = B_A$ by

\[
B_A(f, g, F, G, u, v) := |I_0|^{-1} \sup_{I \subset I \subset I_0} \left| \langle f \rangle_{I_1} - \langle f \rangle_{I_2} \rangle : \langle g \rangle_{I_1} - \langle g \rangle_{I_2} \rangle \cdot |I|,
\]

where the supremum is taken over all (real-valued) functions $f, g$ and the (dyadic) Muckenhoupt weights $w \geq 0$ on $I_0$, such that

\[(3.3)\]

\[
\sup_{I \subset I \subset I_0} \langle w \rangle_I \langle w^{-1} \rangle_I \leq A,
\]

\[(3.4)\]

\[
\langle f \rangle_{I_0} = f, \quad \langle f^2 w \rangle_{I_0} = F, \quad \langle g \rangle_{I_0} = g, \quad \langle g^2 w^{-1} \rangle_{I_0} = G
\]

\[(3.5)\]

\[
\langle w \rangle_{I_0} = u, \quad \langle w^{-1} \rangle_{I_0} = v.
\]

Here again $I_1, I_2$ are the children of the interval $I$.

3.1.1. Properties of $B_A$. The Bellman function $B_A$ satisfies the following properties:

(i) Domain $\text{Dom}(B_A)$ of $B_A$ is given by (3.2): this means that for every set of numbers $f, g, F, G, u, v$ satisfying (3.2) there are functions $f, g$ and a weight $w$ satisfying (3.4), (3.5), so the supremum is well defined (not equal $-\infty$). It also mean, that if the variables are the corresponding averages, they must satisfy the constrains (3.2).

(ii) Range: $0 \leq B_A(f, g, F, G, u, v) \leq CAF^{1/2}G^{1/2}$;

(iii) The main inequality: for any three 6-tuples (let us call them $X = (f, g, F, G, u, v), \quad X_1 = (f_1, g_1, F_1, G_1, u_1, v_1), \quad X_2 = (f_2, g_2, F_2, G_2, u_2, v_2)$) in the domain satisfying

\[
B_A(X) \geq B_A(X_1) + B_A(X_2))/2 + |f_1 - f_2| \cdot |g_1 - g_2|
\]

Note that because of correct homogeneity, the function $B_A$ does not depend on the choice of the interval $I_0$.

Let us explain the properties of the Bellman function $B_A$. The property (i) is easy to explain. Namely, for any weight $w$ and any interval $I$

\[
\langle w \rangle_I \langle w^{-1} \rangle_I \geq 1.
\]

On the other hand the dyadic Muckenhoupt condition $A^2$ means that $\langle w \rangle_I \langle w^{-1} \rangle_I \leq A$ for any $I \in \mathcal{S}$, so the inequalities $1 \leq uv \leq A$ must be satisfied. On the other hand, it is an easy exercise to show that for any $uv$ satisfying $1 \leq uv \leq A$ one can find a dyadic Muckenhoupt weight $w$ satisfying (3.3) and (3.5); one just can consider functions constant on the children of $I_0$.
The inequality $f^2 \leq Fv$ is just the Cauchy–Schwartz inequality:

$$\left|\int_I f \, dx\right| \leq \left(\int_I f^2 w \, dx\right)^{1/2} \left(\int_I w^{-1} \, dx\right)^{1/2}.$$ 

On the other hand, given a weight $w$ it is not hard to find a function $f$ satisfying (3.4): we just put $f = u^{-1/2}f + \phi$, where $\int_{I_0} \phi w \, dx = 0$ and $\int_{I_0} \phi^2 w = F - f^2$.

Property (ii) is pretty straightforward: $B_A(X) \leq 0$ be the definition, and the inequality $B_A(X) \leq CA/F^{1/2}G^{1/2}$ follows from (3.1).

Let us now explain the main inequality (iii). Let $X, X_1, X_2 \in \text{Dom}(B_A)$, $X = (X_1 + X_2)/2$, and let $I_1^0$ and $I_2^0$ be the children of the interval $I_0$. Consider functions $f$ and $g$ and a weight $w$ on $I_0$ such that

$$(3.6) \quad X_{1,2} = ((f)_{I_{1,2}}, (g)_{I_{1,2}}, (f^2 w)_{I_{1,2}}, (g^2 w^{-1})_{I_{1,2}}, (w)_{I_{1,2}}, (w^{-1})_{I_{1,2}})$$

(we just construct the functions on the intervals $I_1^0$ and $I_2^0$ with the prescribed averages there: as we just discussed above such functions always exist). Then $X = (X_1 + X_2)/2 = ((f)_{I_0}, (g)_{I_0}, (f^2 w)_{I_0}, (g^2 w^{-1})_{I_0}, (w)_{I_0}, (w^{-1})_{I_0})$ is the vector of corresponding averages over $I_0$. Denoting by $I_1$ and $I_2$ the children of an interval $I$ we can write

$$|I_0|^{-1} \sum_{I \in \mathcal{I}: I \subset I_0} \left|\langle f \rangle_{I_1} - \langle f \rangle_{I_2}\right| \cdot \left|\langle g \rangle_{I_1} - \langle g \rangle_{I_2}\right| \cdot |I|$$

$$= |f_1 - f_2| \cdot |g_1 - g_2| + |I_0|^{-1} \sum_{I \in \mathcal{I}: I \not\subset I_0} \left|\langle f \rangle_{I_1} - \langle f \rangle_{I_2}\right| \cdot \left|\langle g \rangle_{I_1} - \langle g \rangle_{I_2}\right| \cdot |I|$$

Taking the supremum over all $f, g$ and $w$ satisfying (3.6) we get in the right side $|f_1 - f_2| \cdot |g_1 - g_2| + \left(\mathcal{B}_A(X_1) + \mathcal{B}_A(X_2)\right)/2$.

The supremum in the left side is clearly bounded above by $\mathcal{B}_A(X)$, which proves the main inequality (iii). \hfill \Box

**Remark 3.1.** Notice that the domain of $\mathcal{B}_A$ is not convex, the reason being that the set $\{u, v > 0 : uv \leq A\}$ is not convex.

However, all the other constraints define the convex sets, so the non-convexity of the domain is completely given by the behavior of the coordinates $u, v$. Namely, if the points $X_+, X_- \in \mathbb{R}^6$ are in the domain, and for each $X = X_\theta = \theta X_+ + (1 - \theta)X_-$, $0 < \theta < 1$, we know that $uv \leq A$, then all $X_\theta$, $0 < \theta < 1$, are in the domain.

**Remark 3.2.** The main inequality (iii) implies that the functions $\mathcal{B}_A$ are concave, namely that if the points $X_1, X_2$ and the whole interval $[X_1, X_2] = \{(1 - \theta)X_1 + \theta X_2 : \theta \in [0, 1]\}$ are in $\text{Dom} \mathcal{B}_A$, then for all $\theta \in [0, 1]$

$$(3.7) \quad \mathcal{B}_A((1 - \theta)X_1 + \theta X_2) \geq (1 - \theta)\mathcal{B}_A(X_1) + \theta \mathcal{B}_A(X_2).$$
Condition (iii) implies the so called midpoint concavity, i.e. \((3.7)\) with \(\theta = 1/2\). But that is a well-known fact in convex analysis that for locally bounded concave (convex) functions midpoint concavity (convexity) is equivalent to the regular concavity (convexity), i.e. to the inequality for all \(\theta \in [0,1]\).

The classical reference here would be the monograph \([9]\), see Statement 111 in Section 3.18 where it was stated for convex functions. Note, that only upper bound was assumed in \([9]\), so for the equivalence of midpoint concavity and concavity one can only assume that a function is locally bounded below.

Recall also, that any bounded above convex (respectively bounded below concave) function is continuous, and even locally Lipschitz, see \([4, Theorem 4.1.1]\).

4. The main result: sharp weighted estimates of the Haar shifts

The theorem below is formally the main result of the paper.

Let a dyadic lattice \(\mathcal{D}\) be fixed. Recall that a weight \(w\) satisfies the dyadic Muckenhoupt condition \(A^2_d\) (with respect to the dyadic lattice \(\mathcal{D}\)) if

\[
\sup_{I \in \mathcal{D}} \frac{\langle w \rangle_I}{\langle w^{-1} \rangle_I} =: [w]_{A^2_d} < \infty
\]

**Theorem 4.1** (Main result). Let \(S\) be a Haar shift in \(\mathbb{R}^d\) (in the sense of Definition \([2,2]\)) of complexity \(n\), and let a weight \(w\) satisfies the dyadic Muckenhoupt condition \(A^2_d\). Then the norm of \(S\) in \(L^2(w)\) is at most \(324n2^{2d-2}[w]_{A^2_d}\), where \(C\) is the constant from \((3.1)\).

5. Preliminaries for the proof

5.1. Some simple reductions. Let us simplify the problem a bit.

5.1.1. First slicing. Let us recall that the Haar shift \(S\) was represented as

\[
Sf(x) = \sum_{Q \in \mathcal{D}} |Q|^{-1} \int_Q a_Q(x,y)f(y)dy, \quad \text{supp} a_Q \subset Q \times Q, \|a_Q\|_{\infty} \leq 1,
\]

where \(a_Q\) are given by \((2.3)\).

Let \(n\) be the complexity of the shift \(S\). We can decompose \(S\) as the sum \(S = \sum_{k=0}^{n-1} S_k\), where for \(0 \leq k \leq n\) the operator \(S_k = S_k^n\) is defined by taking the sum only over the cubes \(Q\) of length \(2^{k+nj}, j \in \mathbb{Z}\). If we denote \(\mathcal{L}_k = \mathcal{L}_k^n := \{Q \in \mathcal{D} : \ell(Q) = 2^{k+nj}, j \in \mathbb{Z}\}\), then we can write

\[
S_k f = \sum_{Q \in \mathcal{L}_k} |Q|^{-1} \int_Q a_Q(x,y)f(y)dy
\]

The advantage of the operators \(S_k\) is that they are martingale transforms when we are moving \(n\) units of time at once, so it is possible to apply Bellman function method.
For the readers who are not comfortable thinking in terms of martingales we will just write the estimate

\[
|\langle \mathcal{S}_k f, g \rangle| \leq \sum_{Q \in \mathcal{L}_k} |Q|^{-1} \|\Delta^n_Q f\|_1 \|\Delta^n_Q g\|_1
\]

and notice that the functions $\Delta^n_Q f$ and $\Delta^n_Q g$ are constant on all cubes $R \in \mathcal{L}_k$, $R \subseteq Q$.

5.1.2. Reduction to the real line. First of all it is sufficient to deal only with the Haar shifts on a dyadic lattice in $\mathbb{R}$ (we can assume that we are dealing with the standard dyadic lattice, but the proof for a general dyadic lattice is the same).

So for the slice $\mathcal{S}_k$ of the Haar shift in $\mathbb{R}^d$ we make its representation on the real line by “arranging” cubes along the real line. Namely, for a dyadic cube $Q$ in $\mathbb{R}^d$ take a dyadic interval $I$, $|I| = |Q|$; this interval will correspond to the cube $Q$. Dividing one side of $Q$ into 2, split $Q$ into 2 equal parallelepipeds: then pick a (one to one) correspondence between the parallelepipeds and the children of $I$ (the choice of the correspondence does not matter).

Then by dividing a longer side split each parallelepiped into two equal ones, and and make a correspondence between the new parallelepipeds and the children of the corresponding intervals, see Fig. 1.

After $d$ divisions we end up with the correspondence between the children of $Q$ and the intervals $J \in \text{child}_d(I)$. Note that the intervals $J \in \text{child}_k(I)$, $1 \leq k < d$ correspond to some “almost children” $R$ of $Q$. Here by an “almost child” we mean a parallelepiped some with

![Figure 1. Putting cubes on the line](image-url)
some of the sides coinciding with the sides of \( Q \) and the other sides being halves of the corresponding sides of \( Q \). The construction for \( d = 1 \) is presented on Fig. 1.

One can run this construction up, i.e. for \( \tilde{I} \) being the parent of \( I \) and for \( \tilde{Q} \) being the grandparent of \( Q \) of order \( d \) we can construct, using the procedure described above, a bijection \( \Phi \) from the children and almost children of \( \tilde{Q} \) to the intervals \( J \in \text{chld}_k(\tilde{I}) \), \( 1 \leq k < d \), such that \( \Phi(Q) = I \). To assure that \( \Phi(Q) = I \) one just need at every division make the image of the almost child containing \( Q \) to be the dyadic interval (of the appropriate length) containing \( I \).

A function \( f \in L^1_{\text{loc}}(\mathbb{R}^d) \) will be transferred to a function \( g \in L^1_{\text{loc}}(\mathbb{R}^d) \) such that \( \langle f \rangle_Q = \langle g \rangle_I \), for all \( Q, I, I = \Phi(Q) \).

Let us see what is the “price” to pay for this reduction. First, if \( S \) is a Haar shift (or its slice) of a complexity \( n \) in \( \mathbb{R}^d \), then its model in \( \mathbb{R} \) will be a shift of complexity \( nd \).

The (dyadic) \( A_2 \) norm of the weight on \( \mathbb{R} \) is \( \sup_{R} \langle w \rangle_R \langle w \rangle_R \), where the supremum is taken over all dyadic cubes and over all their almost children. If \( R \) is an almost child of \( Q \), then

\[
\int_{R} w(x) dx \leq \int_{Q} w(x) dx, \quad \int_{R} w(x)^{-1} dx \leq \int_{Q} w(x)^{-1} dx,
\]

and \( |R| \geq 2^{-d+1}|Q| \). So after the transfer to the real line, the Muckenhoupt norm \( \|w\|_{A_2^\infty} \) of the weight increases at most \( 2^{2(d-1)} \) times.

5.2. A technical lemma.

**Lemma 5.1.** Let \( X, X_+, X_- \in \text{Dom} \mathcal{B}_A, X = (X_+ + X_-)/2 \). Then the interval \( [X_-, X_+] \) belongs to \( \text{Dom} \mathcal{B}_{A'} \), where \( A' = 9A/8 \), i.e. for all \( \theta \in (0, 1) \) we have \( X_\theta = \theta X_+ + (1 - \theta)X_- \in \text{Dom} \mathcal{B}_{A'} \).

**Proof.** As it was discussed above in Remark 3.1 we only need to check the non-convex constrain \( uv \leq A \). Because of scaling, it is sufficient to show that if end points and a center of an interval are in the set \( \{(x, y) \in \mathbb{R}^2 : x, y \geq 0, xy \leq 1 \} \), then for any point on the interval \( xy \leq 9/8 \).

It is no hard to see that the worst case scenario is when center and one endpoint is on the line \( y = 1/x \) (and the other endpoint is on a coordinate axis). Since the rescaling \( x \mapsto \alpha x, y \mapsto \alpha^{-1}y, \alpha > 0 \) does not change the domains \( xy \leq C \), we can assume without loss of generality that the center of the interval is at the point \( (1, 1) \). Then the picture for worst case scenario will be as on Fig. 2 or its reflection in the line \( y = x \).

It is an easy exercise to find that the maximal value of \( xy \) on this interval is \( 9/8 \). \( \square \)

6. The main estimate

The following lemma is in the heart of the matter. In fact, it can be considered as the main result of the paper.

**Lemma 6.1** (The main estimate). Let \( \mathcal{B}_A, A > 1 \) be a family of functions satisfying conditions (i), (ii), (iii) from Section 3.1. Let \( I_0 \) be a dyadic interval, and let for all
I \in \text{chld}_k(I_0), \ 0 \leq k \leq n, \ \text{the points } X_I = (f_I, g_I, F_I, G_I, u_I, v_I) \in \text{Dom} B_A \text{ be given. Assume that } X_I \text{ satisfy the dyadic martingale dynamics, namely that if } I_1, I_2 \text{ are the children of } I, \ \text{then}

X_I = \frac{X_{I_1} + X_{I_2}}{2}.

Then for } A' = 4.5A \text{ we have}

\left(2^{-n} \sum_{I \in \text{chld}_n(I_0)} |f_I - f_{I_0}| \right) \left(2^{-n} \sum_{I \in \text{chld}_n(I_0)} |g_I - g_{I_0}| \right)

\leq 72 \left(B_{A'}(X_{I_0}) - 2^{-n} \sum_{I \in \text{chld}_n(I_0)} B_{A'}(X_I) \right).

6.1. Plan of the proof. Let us first explain the idea of the proof informally. We will use the language of random processes (martingales) as the most convenient for the informal explanation, but the formal proof will be completely elementary.

So, we a given a dyadic martingale } X_I, \ I \in \mathcal{D}, \ \text{where at each point we have 2 choices with equal probability, and we want to estimate below the difference between } B_{A'}(X_I) \text{ at the initial moment } I = I_0 \text{ and its expected value after } n \text{ steps.}

An immediate idea would be to use the property (iii) of the Bellman function to estimate expected loss at each step. But this would not work: the conditional expectation of the loss when going from the state } X_I \text{ to the states } X_{I_1}, X_{I_2}, \ \text{where } I_{1,2} \text{ are the children of } I
is estimated below by \( c|f_I - f_J| \cdot |g_I - g_J| \), and this quantity can be small, even can be zero: imagine a situation when at each step only one of the variables \( f \) or \( g \) is changed.

So, the main idea is to change the martingale, preserving the starting point \( X_{I_0} \) and the final values and probabilities.

Namely, from the starting point \( X_{I_0} \) we move with probabilities 1/2 to the points \( X^+ \) and \( X^- \), and \( X^\pm = X_{I_0}^\pm = (f^\pm, g^\pm, F^\pm, G^\pm, u^\pm, v^\pm) \), \( X_{I_0} = (X^+ + X^-)/2 \). We will move “sufficiently far”, so \( |f^\pm - f_{I_0}| \) and \( |g^\pm - g_{I_0}| \) are comparable to \( 2^{-n} \sum_{I \in \text{child}_n(I_0)} |f_I - f_{I_0}| \) and \( 2^{-n} \sum_{I \in \text{child}_n(I_0)} |g_I - g_{I_0}| \) respectively, but “not too far”, so the points \( X^\pm \) will be “almost averages” of the points \( X_I, I \in \text{child}_n(I_0) \).

Since we moved sufficiently far, we get the amount we need at the first step. Moreover, since the points \( X^\pm \) are “almost averages” of the points \( X_I, I \in \text{child}_n(I_0) \), we can start the process from the points \( X^\pm \) to end up after \( n \) steps at the points \( X_I, I \in \text{child}_n(I_0) \).

Of course, it will not be possible to get from \( X^+ \) (or \( X^- \)) to the end points \( X_I, I \in \text{child}_n(I_0) \) via the standard dyadic martingale, when at each point we have 2 choices with equal probability. But since the first step was sufficiently small, it will still be possible to get to the desired endpoints via a binary martingale, where at each point we have 2 choices with not equal but with almost equal probability.

It might be intuitively clear from the symmetry of \( X^\pm \), that after moving from \( X_{I_0} \) to \( X^\pm \) it is possible to start the martingale from \( X^\pm \) so we get from \( X_{I_0} \) to the endpoints with equal probability. More precisely, we will get from \( X^+ \) to \( X_I \), \( I \in \text{child}_n(I_0) \) with probability \( 2^{-n}(1 + \alpha_I) \), \( |\alpha_I| \leq 1/3 \), and we get from \( I^- \) to the same point \( I \) with the probability \( 2^{-n}(1 - \alpha_I) \), so the probability of getting from \( I_0 \) to \( I \) will be \( 2^{-n} \).

If this reasoning is not intuitively clear to the reader, he or she should not worry, because the rigorous proof (not requiring any probability) will be presented later; the probabilistic interpretation will guide us through the formal calculations.

### 6.2. Formal proof: another technical lemma.

Let us recall a theorem by K. Ball [2 Theorem 7], solving the famous Tarski plank problem for convex bodies. This theorem also can be treated as a “multiple Hahn–Banach Theorem”.

**Theorem 6.2.** Let \( x_k \) be unit vectors \( \|x_k\| = 1 \) in a (real) normed space \( \mathcal{X} \), and let \( m_k \in \mathbb{R}, w_k \geq 0 \) such that

\[
\sum_k w_k = 1.
\]

Then there exist a functional \( x^* \in \mathcal{X}^* \), \( \|x^*\| \leq 1 \) such that

\[
|\langle x_k, x^* \rangle - m_k| \geq w_k \quad \text{for all} \ k.
\]

Note that the number of vectors here can be infinite.

This theorem with 2 vectors, \( m_k = 0, w_1 = w_2 = 1/2 \) gives as the following simple lemma that we will use.
Lemma 6.3. Let $X$ be a (real) normed space, and let $a, b \in X$, $\|a\| = \|b\| = 1$. There exists $x^* \in X^*$, $\|x^*\| = 1$ such that

$$|\langle a, x^* \rangle| \geq \frac{1}{2}, \quad |\langle b, x^* \rangle| \geq \frac{1}{2}.$$  

Note, that for our purposes we just need this lemma with some constant, not necessarily with the optimal one $1/2$; we just get a worse constant in Theorem 4.1. A proof of this lemma with some constant is an easy exercise in elementary functional analysis.

6.3. Formal proof: the first step. Let $N = 2^n$, and let $\ell^p_N$ be the space $\mathbb{R}^N$ endowed with the $\ell^p$ norm. Define $e \in \ell^p_N$, $e = (1, 1, \ldots, 1)$. Consider the quotient space $X = \ell^1_N/\text{span}\{e\}$. For $x \in \ell^1_N$ let $x^0 := x - N^{-1}\langle x, e \rangle e$.

Then

$$\|x\|_X \leq \|x^0\|_{\ell^1_N} \leq 2\|x\|_X. \tag{6.1}$$

Indeed, the first inequality is trivial (follows from the definition of the norm in the quotient space). As for the second one, $|\langle x, e \rangle| \leq \|x\|_{\ell^1_N}$, $\|e\|_{\ell^1_N} = N$, so it follows from the triangle inequality that

$$\|x^0\|_{\ell^1_N} \leq \|x\|_{\ell^1_N} + N^{-1}|\langle x, e \rangle|\|e\|_{\ell^1_N} \leq 2\|x\|_{\ell^1_N}.$$  

This inequality remains true if one replaces $x$ by $x - \alpha e$, $\alpha \in \mathbb{R}$, so the second inequality in (6.1) is proved.

The dual space $X^*$ can be identified with a subspace of $\ell^\infty_N$ consisting of $x^* \in \ell^\infty_N$ such that $\langle e, x^* \rangle = 0$ (with the usual $\ell^\infty_N$-norm).

Applying Lemma 6.3 to the above space $X$ with $a$ and $b$ being the normalized vectors $\{f_I - f_{I_0}\}_{I \in \text{child}_n(I_0)}$ and $\{g_I - g_{I_0}\}_{I \in \text{child}_n(I_0)}$ respectively, we get that there exists $x^* = \{\alpha_I\}_{I \in \text{child}(I_0)}$ such that

$$\sum_{I \in \text{child}_n(I_0)} \alpha_I = 0, \tag{6.2}$$

$$|\alpha_I| \leq \frac{1}{3} \quad \forall I \in \text{child}_n(I_0), \tag{6.3}$$

and

$$\left| \sum_{I \in \text{child}_n(I_0)} \alpha_I (f_I - f_{I_0}) \right| \geq \frac{1}{12} \sum_{I \in \text{child}_n(I_0)} |f_I - f_{I_0}|, \tag{6.4}$$

$$\left| \sum_{I \in \text{child}_n(I_0)} \alpha_I (g_I - g_{I_0}) \right| \geq \frac{1}{12} \sum_{I \in \text{child}_n(I_0)} |g_I - g_{I_0}|. \tag{6.5}$$

Note that Lemma 6.3 gives the comparison with the norm in $X = \ell^1_N/\text{span} e$ with the constant $1/6$, and the second inequality in (6.1) give as an extra factor $1/2$.

Below, all identities (inequalities) with $\pm$ mean that there are two sets of identities (inequalities): one with $+$ and the other with $-$. 
Define $X^\pm = X^\pm_{I_0} = (f^\pm, g^\pm, F^\pm, G^\pm, u^\pm, v^\pm)$ by

$$X^\pm = 2^{-n} \sum_{I \in \text{child}_n(I_0)} (1 + \alpha_I) X_I = X_{I_0}^\pm + 2^{-n} \sum_{I \in \text{child}_n(I_0)} \alpha_I X_I,$$

so $X_{I_0} = (X^+ + X^-)/2$.

Note that (6.4), (6.5) and the identities

$$f_{I_0} = 2^{-n} \sum_{I \in \text{child}_n(I_0)} f_I, \quad g_{I_0} = 2^{-n} \sum_{I \in \text{child}_n(I_0)} g_I$$

imply that

$$|f^\pm - f_{I_0}| \geq \frac{1}{12} 2^{-n} \sum_{I \in \text{child}_n(I_0)} |f_I - f_{I_0}|,$$

$$|g^\pm - g_{I_0}| \geq \frac{1}{12} 2^{-n} \sum_{I \in \text{child}_n(I_0)} |g_I - g_{I_0}|.$$ 

Let

$$x_I^\pm := 1 \pm \alpha_I;$$

note that $2/3 \leq x_I \leq 4/3$.

6.4. **Formal proof: modifying the martingale.** For $I \in \text{child}_k(I_0)$, $1 \leq k < n$ define

$$X_I^\pm := \left( \sum_{J \in \text{child}_n(I_0): J \subset I} x_J^\pm X_J \right) \div \left( \sum_{J \in \text{child}_n(I_0): J \subset I} x_J^\pm \right).$$

Note, that for $I \in \text{child}_n(I_0)$ we have $X_I^+ = X_I^- = X_I$, where $X_I$ are the points given in the statement of the lemma.

For $I \in \text{child}(K)$ define

$$\theta_I^\pm := \left( \sum_{J \in \text{child}_n(I_0): J \subset I} x_J^\pm \right) \div \left( \sum_{J \in \text{child}_n(I_0): J \subset K} x_J^\pm \right).$$

Note that $\theta_I^\pm \geq 0$, and it $I_1$, $I_2$ are the children of $I$, then

$$\theta_{I_1}^\pm + \theta_{I_2}^\pm = 1, \quad \theta_{I_1}^\pm X_{I_1}^\pm + \theta_{I_2}^\pm X_{I_2}^\pm = X_I^\pm,$$

i.e. $X_{I_1}^+$ (respectively $X_{I_2}^-$) is in the interval connecting $X_{I_1}^+$ and $X_{I_2}^+$ (respectively $X_{I_1}^-$ and $X_{I_2}^-$). From the probabilistic point of view, $\theta_{I_1}^\pm$ and $\theta_{I_2}^\pm$ are the probabilities of moving from $X_I^\pm$ to $X_{I_1}^\pm$ and $X_{I_2}^\pm$, respectively.

It is not hard to show that $\theta_{I_1, I_2}^\pm$ cannot be too close to 0 or 1, but we do not need this fact for the formal proof.
Identity (6.6) means that the points $X^\pm_I$ are in the convex hull of the points $X_J : J \in \text{chld}_n(I_0), J \subset I$. We claim that moreover, $X^\pm_I \in \text{Dom}(B_{4A})$ for all $I \in \text{chld}_k(I_0)$, $0 \leq k \leq n$.

Indeed, since $X^\pm_I$ are in the convex hull of the points $X_J \in \text{Dom}(B_A) \subset \text{Dom}B_{4A}$, and among the constrains defining $\text{Dom}(B_{4A})$ only the constrain

$$uv \leq 4A$$

is not convex, we only need to check this constrain.

Notice that equation (6.6) with all $x^\pm_I$ replaced by 1 gives us $X_I$, which belongs to $\text{Dom}(B_A)$. Let us look at the $u$-coordinate of $X^\pm_I$: we can easily see that

$$u^\pm_I \leq u_I \frac{4/3}{2} = 2u_I,$$

because the maximal possible numerator in the $u$-coordinate of (6.6) happens when all $x^\pm_I = 4/3$ and the minimal possible numerator when all $x^\pm_I = 2/3$. The same holds for $v^\pm_I$, so $u^\pm_I v^\pm_I \leq 4u_I v_I \leq 4A$, thus $X^\pm_I \in \text{Dom}(B_{4A})$.

We can also show that if $I_1$ and $I_2$ are the children of $I$, then the centers of the intervals $[X^\pm_{I_1}, X^\pm_{I_2}]$ also belong to $B_{4A}$. Indeed, we know that $X_I = (X_{I_1} + X_{I_2})/2$ (the center of $[X^\pm_{I_1}, X^\pm_{I_2}]$) belongs to $B_A$. As we discussed above,

$$u^\pm_{I_{1,2}} \leq 2u_{I_{1,2}}, \quad v^\pm_{I_{1,2}} \leq 2v_{I_{1,2}}.$$

Therefore, if $\bar{u}^\pm$ and $\bar{v}^\pm$ are the $u$ and $v$ coordinates of the centers of the intervals $[X^\pm_{I_1}, X^\pm_{I_2}]$, we can conclude that

$$\bar{u}^\pm \leq 2u_I, \quad \bar{v}^\pm \leq 2v_I$$

thus the centers of the intervals $[X^\pm_{I_1}, X^\pm_{I_2}]$ also belong to $B_{4A}$ (as we discussed above, we only need to check the non-convex constrain $uv \leq 4A$).

Therefore, by Lemma 5.1 the intervals $[X^\pm_{I_1}, X^\pm_{I_2}]$ are in $B_{A'}$, where $A' = 4A(9/8) = 4.5A$.

### 6.5. Conclusion of the proof.

Now we are ready to complete the proof. By property (iii) of the Bellman function, using the fact that $|f^+ - f^-| = 2|f^\pm - f_{I_0}|$,

$$|f^\pm - f_{I_0}| \cdot |g^\pm - g_{I_0}| \leq \frac{1}{4} \left(B_{A'}(X_{I_0}) - \frac{1}{2} (B_{A'}(X^+) + B_{A'}(X^-))\right)$$

(6.9)

It follows from the concavity of $B_{A'}$ and (6.8) that if $I_1$ and $I_2$ are the children of $I$, then

$$B_{A'}(X^\pm_I) \geq \theta_{I_1} B_{A'}(X^\pm_{I_2}) + \theta_{I_2} B_{A'}(X^\pm_{I_2});$$

recall that as we discussed above, the whole intervals $[X^\pm_{I_1}, X^\pm_{I_2}]$ are in $\text{Dom}(B_{8A})$. 

SERGEI TREIL
Let us apply this inequality to $I_0$, then substitute in the right side the inequalities for the children of $I_0$, and so on. Then, using the fact that for $I \in \text{chld}_n(I_0)$

$$\prod_{J \in \mathcal{D}: I \subset J \supseteq I_0} \theta_J^\pm = x_I^\pm \div \left( \sum_{J \in \text{chld}_n(I_0)} x_J^\pm \right) = x_I^\pm 2^{-n},$$

we get the inequality

$$B_{A'}(X^\pm) \geq 2^{-n} \sum_{I \in \text{chld}_n(I_0)} x_I^\pm B_{A'}(X_I);$$

note that we are using $X_I$ instead of $X_I^\pm$ in the right side, because $X_I^+ = X_I^- = X_I$ for $I \in \text{chld}_n(I_0)$.

Substituting this inequality into (6.9) and taking into account that $x_I^+ + x_I^- = 2$ for $I \in \text{chld}_n(I_0)$, we get

$$|f^\pm - f_{I_0}| \cdot |g^\pm - g_{I_0}| \leq \frac{1}{4} \left( B_{A'}(X_{I_0}) - 2^{-n} \sum_{I \in \text{chld}_n(I_0)} B_{A'}(X_I) \right)$$

(6.10)

The estimates (6.4) and (6.5) mean that

$$2^{-n} \sum_{I \in \text{chld}_n(I_0)} |f_I - f_{I_0}| \leq 12 |f^\pm - f_{I_0}|,$$

(6.11)

$$2^{-n} \sum_{I \in \text{chld}_n(I_0)} |g_I - g_{I_0}| \leq 12 |g^\pm - g_{I_0}|.$$

(6.12)

Combining these estimate with (6.10), we get the conclusion of the lemma. □

7. Proof of the main result (Theorem 4.1)

As we discussed above, we need to estimate the “slices” $S_k$ of the Haar shift, or equivalently, their models on the real line $\mathbb{R}$. Let $S = S_k$ be such a model of a slice, and let $n$ be its complexity. As it was shown before

$$|\langle S_k f, g \rangle| \leq \sum_{I \in \mathcal{L}_k} |I|^{-1} \|\Delta_J^n f\|_1 \|\Delta_J^n g\|_1$$

For $I \in \mathcal{D}$ let $X_I := (\langle f \rangle_I, \langle g \rangle_I, \langle f w^{-1} \rangle_I, \langle g w^{-1} \rangle_I, \langle w \rangle_I, \langle w^{-1} \rangle_I)$. Define $A := [w]_{A_2}$, so $X_I \in \text{Dom}(B_{A'})$ for all $I \in \mathcal{D}$. Lemma 6.1 states that

$$|I|^{-1} \|\Delta_J^n f\|_1 \|\Delta_J^n g\|_1 \leq |I|B_{A'}(X_I) - \sum_{J \in \text{chld}_n(I)} |J|B_{A'}(X_J)$$

\[\square\]
Writing this estimate for each $J \in \text{chld}_n(I)$, then repeating this $m$ times we get

$$
\sum_{J \in \mathcal{L}_k: J \subseteq I \atop \ell(J) > 2^{-nm} \ell(I)} |J|^{-1} \|\Delta^n_J f\|_1 \|\Delta^n_J g\|_1 \leq 72 \left( |I| B_{A'}(X_I) - \sum_{J \in \text{chld}_m(I)} |J| B_{A'}(X_J) \right)
$$

$$
\leq 4.5 \cdot 72 \cdot CA \left( f^2 w \right)^{1/2} \left( g^2 w^{-1} \right)^{1/2} |I| \leq 324CA \|f\|_{L^2(w)} \|g\|_{L^2(w^{-1})}
$$

where $C$ is the constant from the property (ii) of the Bellman functions $B_{A'}$; here the second inequality holds because $B_{A'}(X_I) \leq CA' F^{1/2}_I G^{1/2}_I = 4.5CA' F^{1/2}_I G^{1/2}_I$ and $B_{A'}(X_J) \geq 0$ by property (ii) of the Bellman function.

Letting $m \to \infty$ we get

$$
\sum_{J \in \mathcal{L}_k: J \subseteq I} |J|^{-1} \|\Delta^n_J f\|_1 \|\Delta^n_J g\|_1 \leq 324CA \|f\|_{L^2(w)} \|g\|_{L^2(w^{-1})}
$$

Covering the line by the intervals $I \in \mathcal{L}$ of length $2^M$ and applying the above inequality to each $I$ we get

$$
\sum_{J \in \mathcal{L}_k: |J| \leq 2^M} |J|^{-1} \|\Delta^n_J f\|_1 \|\Delta^n_J g\|_1 \leq 324CA \|f\|_{L^2(w)} \|g\|_{L^2(w^{-1})},
$$

and letting $M \to \infty$ we get that the norm of each slice $\mathcal{S}_k$ is bounded by $324CA$. Recall, that we had $n$ slices. Recall that $A$ is the $A_2$ norm of the weight $w$ transferred to the real line $\mathbb{R}$. As it was discussed above in Section 5.1.2, it is estimated by $2^{2d-2}[w]_{A_2}$, where $[w]_{A_2}$ is the $A_2$ norm of the original weight $w$ in $\mathbb{R}^d$.

Gathering everything together we get the conclusion of the main result (Theorem 4.1).

8. Estimates of the paraproducts

As it was mentioned before in the Introduction, to prove the $A_2$ conjecture for general Calderón–Zygmund operators, besides getting linear in $[w]_{A_2}^{d}$ and subexponential in complexity estimate for the Haar shifts, one also need a linear in $[w]_{A_2}^{d}$ estimate of the so-called paraproducts. Since the paraproducts have a fixed complexity 1, one does not care about growth of the estimates with complexity, and any (linear in $[w]_{A_2}$) estimate of the paraproduct, for example one obtained in [3] would work.

Here we would like to show how using Bellman function approach to get linear in $[w]_{A_2}^{d}$ estimate of a general paraprodut from the estimate for the simplest paraproduct in $\mathbb{R}$ obtained by O. Beznosova [3].

Let us recall the main definitions. Let a dyadic lattice $\mathcal{D}$ in $\mathbb{R}^d$ be fixed.

**Definition 8.1.** Let $\varphi$ be a locally integrable function. A paraproduct $\Pi_\varphi$ with symbol $\varphi$ is defined by

$$
\Pi_\varphi f := \sum_{I \in \mathcal{D}} \langle f \rangle_I \Delta_I \varphi.
$$
It is well known that the paraproduct $\Pi_{\varphi}$ is bounded in $L^2$ (unweighted) if and only if $\varphi \in \text{BMO}^d$, i.e. if and only if
\[
\sup_{J \in \mathcal{D}} \frac{1}{|J|} \sum_{I \in \mathcal{D}: I \subset J} \|\Delta_I \varphi\|_2^2 =: \|\varphi\|_{\text{BMO}^d}^2 < \infty.
\]
This statement is in fact equivalent to the dyadic Carleson Embedding Theorem, and from the sharp estimates in the Embedding Theorem one can get that norm of $\Pi_{\varphi}$ is bounded by $2\|\varphi\|_{\text{BMO}^d}$.

As for the linear in $[w]_{A_2}^d$ weighted estimates of the paraproducts, it was proved by O. Beznosova in [3] that for the paraproduct on the real line
\[
\|\Pi_{\varphi}\|_{L^2(w) \to L^2(w)} \leq C_1 \|f\|_{\text{BMO}^d} [w]_{A_2}^d.
\]
where $C_1$ is an absolute constant.

This result can be rewritten as
\[
(8.1) \quad \sum_{I \in \mathcal{D}} \langle f \rangle_I \|\Delta_I \varphi\|_2 \langle g \rangle_{I_1} - \langle g \rangle_{I_1} |I|^{1/2} \leq C [w]_{A_2}^d \|\varphi\|_{\text{BMO}^d} \|f\|_{L^2(w)} \|g\|_{L^2(w^{-1})}
\]
for all $f \in L^2(w)$ and all $g \in L^2(w^{-1})$ (because changing the signs in the decomposition $\varphi = \sum_{I \in \mathcal{D}} \Delta_I \varphi$ does not change $\|\varphi\|_{\text{BMO}^d}$). Here $\langle \cdot \rangle$ is the unweighted $L^2$ norm, $I_1$ and $I_2$ denote the children of $I$ and $C = C_1/2$.

We want to extend this result to $\mathbb{R}^d$. We will show how to prove the following proposition

**Proposition 8.2.** Let $\varphi \in \text{BMO}^d(\mathbb{R}^d)$, and let $\Pi_{\varphi}$ be the corresponding paraproduct in $L^2(\mathbb{R}^d)$. Let a weigh $w$ in $\mathbb{R}^d$ satisfies the dyadic Muckenhoupt condition $A_2^d$. Then $\Pi_{\varphi}$ is bounded in $L^2(w)$ with the norm at most $16222^{d-2}C_2 [w]_{A_2}^d$, where $C_2$ is the constant from (8.1).

We present the proof of this result using the Bellman function from the result for the real line. Note that because of homogeneity it is sufficient to prove Proposition 8.2 only for $\|\varphi\|_{\text{BMO}^d} \leq 1$.

### 8.1. Bellman function for paraproducts on the real line.

For an interval $I_0 \in \mathcal{D}$ let us fix the quantities
\[
(8.2) \quad \langle f \rangle_{I_0} = f, \quad \langle f^2 w \rangle_{I_0} = F, \quad \langle g \rangle_{I_0} = g, \quad \langle g^2 w^{-1} \rangle_{I_0} = G
\]
\[
(8.3) \quad \langle w \rangle_{I_0} = u, \quad \langle w^{-1} \rangle_{I_0} = v.
\]
\[
(8.4) \quad M = |I_0|^{-1} \sum_{I \in \mathcal{D}: I \subset I_0} \|\Delta_I \varphi\|_2^2
\]
and define the function $B_A(f, g, F, G, u, v, M)$ as
\[
B_A(f, g, F, G, u, v, M) = |I_0|^{-1} \sup_{I \in \mathcal{D}: I \subset I_0} \sum_{I \in \mathcal{D}: I \subset I_0} \langle f \rangle_I \|\Delta_I \varphi\|_2 \langle g \rangle_{I_1} - \langle g \rangle_{I_1} |I|^{1/2}
\]
where the supremum is taken over all \( f \in L^2(w) \), \( g \in L^2(w^{-1}) \), all weights \( w \), \( |w|_{A_2}^d \leq A \) and all \( \varphi \in \text{BMO}^d \), \( \|\varphi\|_{\text{BMO}^d} \leq 1 \), satisfying (8.2)–(8.4).

8.1.1. Properties of Bellman function.

(i) Domain \( \text{Dom}(B_A) \) is given by the conditions

\[
\begin{align*}
u, v > 0, \quad 1 &\leq uv \leq A, \quad f^2 \leq Fv, \quad g^2 \leq Gu, \quad 0 \leq M \leq 1;
\end{align*}
\]

this simply means that for any choice of the appropriate functions the corresponding averages satisfy these constrains and that for any 7-tuple of reals satisfying these constrains there are functions with corresponding averages over \( I_0 \), so the supremum is well defined (not \(-\infty\)).

(ii) Range:

\[
0 \leq B(X) \leq CA^{1/2}F^{1/2}G^{1/2}.
\]

(iii) The main inequality: Let \( X = (f, g, F, G, u, v) \). Then for all

\[
(X, M), (X_{1/2}, M_{1/2}) \in \text{Dom}(B_A),
\]

such that \( X = (X_1 + X_2)/2, M - (M_1 + M_2)/2 =: d \geq 0 \), the following inequality holds:

\[
B_A(X, M) - (B_A(X_1, M_1) + B_A(X_2, M_2))/2 \geq d|f| \cdot |g_1 - g_2|.
\]

The proof of the properties of the Bellman function is pretty standard, one can do it following the lines of Section 3.1.1.

8.2. The main estimate. The proof of Proposition 8.2 follows easily from the lemma below. The details of the reduction are essentially the same as for the Haar shifts (slicing, remodeling on the real line and then estimating each slice), so we leave it as an easy exercise for the reader.

Lemma 8.3 (The main estimate). Let \( B_A, A > 1 \) be a family of functions satisfying conditions (i), (ii), (iii) from Section 8.1.1. Let \( I_0 \) be a dyadic interval, and let for all \( I \in \text{chld}_k(I_0), 0 \leq k \leq n \), the points \( X_I = (f_I, g_I, F_I, G_I, u_I, v_I) \), \( M_I, (X_I \in \text{Dom} B_A \) be given. Assume that \( X_I \) satisfy the dyadic martingale dynamics, namely that if \( I_1, I_2 \) are the children of \( I \), then

\[
X_I = \left( X_{I_1} + X_{I_2} \right)/2.
\]

Assume also that \( M_I = (M_{I_1} + M_{I_2})/2 \) for all \( I \) except \( I_0 \), and let

\[
M_{I_0} - 2^{-n} \sum_{I \in \text{chld}_n(I_0)} M_I =: d_{I_0} \geq 0.
\]
Then for $A' = 4.5A$ we have

$$d_{I_0} |f_{I_0}| \left( 2^{-n} \sum_{I \in \text{chld}_n(I_0)} |g_I - g_{I_0}| \right)$$

$$\leq 36 \left( B_{A'}(X_{I_0}, M_{I_0}) - 2^{-n} \sum_{I \in \text{chld}_n(I_0)} B_{A'}(X_I, M_I) \right).$$

To prove this lemma one can just literally follow the proof of Lemma 6.1. The main difference (a simplification) is that when we are picking weights $\alpha_I, |\alpha_I| \leq 1/3$ for the first step, we only need to get the estimate (6.5), which can be done trivially (because we only need to control one parameter) and with the constant $1/6$ instead of $1/12$. This give the constant 36 instead of 72 in the conclusion of the lemma.

Then we define

$$(X^\pm, M^\pm) = (X^\pm_{I_0}, M^\pm_{I_0}) = 2^{-n} \sum_{I \in \text{chld}_n(I_0)} (X_I, M_I),$$

and the rest of the proof is the same as for Lemma 6.1 with $X^\pm_I$ replaced by $(X^\pm_I, M^\pm_I)$. □

9. Concluding remarks

The heart of this paper is definitely Lemma 6.1 which is a very general fact about convex function. Concave functions with concavity estimated by $|\Delta f| \cdot |\Delta g|$ are very common in the dyadic Harmonic analysis: they appear if one writes the estimates of the bilinear form of dyadic martingale multipliers $T_\sigma$.

Lemma 6.1 gives a simple way to transfer the estimates from the dyadic martingale multipliers to more general martingale transforms, like dyadic shifts, etc. As the story of the $A_2$ conjecture illustrates, such transference until now was considered highly non-trivial, even in the case of simplest possible non-multiplier martingale transforms (like the simplest Haar shift considered by S. Petermichl).

It also probably worth mentioning that for the second degree polynomials an analogue of Lemma 6.1 was known for some time. Namely, it is a known fact (cf. [6]) that if $Q[\ldots, x, y, \ldots]$ is a quadratic form such that

$$Q[\ldots, x, y, \ldots] \geq 2|xy|,$$

then there exists $\alpha > 0$ such that

$$Q[\ldots, x, y, \ldots] \geq \alpha x^2 + \alpha^{-1} y^2.$$
Therefore, if $\Phi$ is a second degree polynomial of variables $X = (\ldots, x, y, \ldots)$ such that for $X = (X_1 + X_2)/2$

$$\Phi(X) - (\Phi(X_1) + \Phi(X_2))/2 \geq |x_1 - x_2| \cdot |y_1 - y_2|$$

then for any $\theta_k \geq 0$, $\sum_k \theta_k = 1$ and for any $X, X_k$ such that $X = \sum_k \theta_k X_k$,

$$\Phi(X) - \sum_k \theta_k \Phi(X_k) \geq 4 \left( \sum_k \theta_k |x - x_k|^2 \right)^{1/2} \left( \sum_k \theta_k |y - y_k|^2 \right)^{1/2}$$

If we take $2^n$ terms and put $\theta_k = 2^{-n}$ we get (for the second degree polynomials) a stronger version of Lemma 6.1 where we estimate above not the product of $L^1$ norms, but the bigger product of $L^2$ norms.

References

[1] K. Astala, T. Ivaniec, E. Saksman, *Beltrami operators in the plane*, Duke Math J., 107 (2001), 27-56.
[2] K. Ball, *The plank problem for symmetric bodies*, Invent. Math. 104 (1991), no. 3, 535–543.
[3] O. Beznosova, *Linear bound for the dyadic paraproduct on weighted Lebesgue space $L^2(w)$*, J. Funct. Analysis, 255 (2008), No. 4, 994–1007.
[4] J. M. Borwein and A. S. Lewis, *Convex analysis and nonlinear optimization*, second ed., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3, Springer, New York, 2006, Theory and examples.
[5] S. M. Buckley, *Estimates for operator norms on weighted spaces and reverse Jensen inequalities*, Trans. Amer. Math. Soc., 340 (1993), no. 1, p53–272.
[6] O. Dragičević, S. Treil, and A. Volberg, *A theorem about three quadratic forms*, Int. Math. Res. Not. IMRN (2008), Art. ID run 072, 9pp.
[7] O. Dragičević and A. Volberg, *Sharp estimate of the Ahlfors-Beurling operator via averaging martingale transforms*, Michigan Math. J. 51 (2003), no. 2, 415–435.
[8] T. Figiel, *Singular integral operators: a martingale approach*, Geometry of Banach spaces (Strobl, 1989), London Math. Soc. Lecture Note Ser., vol. 158, Cambridge Univ. Press, Cambridge, 1990, pp. 95–110.
[9] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, at the University Press, 1952, 2d ed.
[10] T. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, arXiv:1007.4330.
[11] T. Hytönen, C. Pérez, S. Treil, and A. Volberg, *Sharp weighted estimates of the dyadic shifts and $A_2$ conjecture*, arXiv:1010.0755v1 [math.CA] (2010), 36pp.
[12] R. Hunt, B. Muckenhoupt, R. Wheeden, *Weighted norm inequalities for the conjugate function and the Hilbert transform*, Trans. Amer. Math. Soc., 176 (1973), pp. 227-251.
[13] M. Lacey, S. Petermichl, M. Reguera, *Sharp $A_2$ inequality for Haar shift operators*, Math. Ann., 348 (2010), 127–141.
[14] F. L. Nazarov and S. R. Treil, *The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis*, Algebra i Analiz 8 (1996), no. 5, 32–162.
[15] F. Nazarov, S. Treil and A. Volberg, *Bellman function in stochastic control and harmonic analysis*, Systems, approximation, singular integral operators, and related topics (Bordeaux, 2000), 393–423, Oper. Theory Adv. Appl., 129, Birkhäuser, Basel, 2001.
[16] F. Nazarov, S. Treil, and A. Volberg, *The Bellman functions and two-weight inequalities for Haar multipliers*, J. of Amer. Math. Soc., 12, (1999), no. 4, 909-928.
[17] F. Nazarov, S. Treil, and A. Volberg, *The Tb theorem on non-homogeneous spaces*, Acta Math., 190 (2003), 151–239.

[18] F. Nazarov, S. Treil, and A. Volberg, *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Res. Lett. 15 (2008), no. 3, 583–597.

[19] C. Pérez, S. Treil, A. Volberg, *On $A_2$ conjecture and corona decomposition of weights*, arxiv1005.2630.

[20] 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 (2007), no. 5, 1355–1375.

[21] S. Petermichl, *The sharp weighted bound for the Riesz transforms*. Proc. Amer. Math. Soc. 136 (2008), no. 4, 1237–1249.

[22] S. Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*. C. R. Acad. Sci. Paris Sér. I Math. 330 (2000), no. 6, 455–460.

[23] S. Petermichl, A. Volberg, *Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular*, Duke Math. J., 112 (2002), no. 2, 281-305.

[24] A. Reznikov, S. Treil, and A. Volberg, *A sharp estimate of weighted dyadic shifts of complexity 0 and 1*, arXiv:1104.5347v1 [math.CA] (2011), 8pp.

[25] E. Sawyer, *A characterization of a two-weight norm inequality for maximal operators*, Studia Math., 75 (1982), no. 1, pp. 1-11.

[26] A. Vagharshakyan, *Recovering singular integral kernels from Haar shifts*, Proc. Amer. Math. Soc., 138 (2010), 4303–4309.

[27] J. Wittwer, *A sharp estimate on the norm of the martingale transform*. Math. Res. Lett. 7 (2000), no. 1, 1–12.

Dept. of Mathematics, Brown University, 151 Thayer Str./Box 1917, Providence, RI 02912, USA

E-mail address: treil@math.brown.edu

URL: http://www.math.brown.edu/~treil