ON THE FAILURE OF LOWER SQUARE FUNCTION ESTIMATES IN THE NON-HOMOGENEOUS WEIGHTED SETTING

Komla Domelevo, P Ivanisvili, Stefanie Petermichl, S Treil, A Volberg

ON THE FAILURE OF LOWER SQUARE FUNCTION ESTIMATES IN THE NON-HOMOGENEOUS WEIGHTED SETTING

K. DOMELEVO, P. IVANISVILI, S. PETERMICHL, S. TREIL, AND A. VOLBERG

ABSTRACT. We show that the classical A_∞ condition is not sufficient for a lower square function estimate in the non-homogeneous weighted L^2 space. We also show that under the martingale A_2 condition, an estimate holds true, but the optimal power of the characteristic jumps from $1/2$ to 1 even when considering the classical A_2 characteristic. This is in a sharp contrast to known estimates in the dyadic homogeneous setting as well as the recent positive results in this direction on the discrete time non-homogeneous martingale transforms. Last, we give a sharp A_∞ estimate for the n-adic homogeneous case, growing with n.

CONTENTS

1. Introduction 1
2. Setup and motivations 3
3. Main results 6
4. Reduction of lower bound to an embedding theorem 8
5. Counterexample for the A_2 lower bound. 12
6. No bounds in terms of A_∞ 19
7. Estimate in terms of martingale $A_\infty^{\mathcal{D}}$ for homogeneous filtrations 19
8. Upper bound for the square function 23
References 25

1. INTRODUCTION

It is a classical result that the Haar system on the real line is an unconditional basis in the weighted space $L^2(w) = L^2(\mathbb{R}, w)$ if and only if the weight w satisfies the dyadic Muckenhoupt A_2 condition. This is equivalent to boundedness of the predictable ± 1 multiplier on the martingale difference sequences with underlying homogeneous dyadic filtration. This generalizes to martingale difference spaces in homogeneous filtrations. These results were proved in [16], where also Littlewood–Paley estimates were considered. It has been known for some time that the optimal unconditional basis

Komla Domelevo and Stefanie Petermichl were supported by ERC project CHRiSHarMa no. DLV-682402. Sergei Treil was supported by the NSF grants DMS-1600139. Alexander Volberg was supported by NSF grant DMS-1600065. The paper was written while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring semester of 2017, supported by the National Science Foundation under Grant No. 1440140.

constants are the first power of the A_2 characteristic of the weight. Through averaging, it follows that the square function has no worse upper bounds, so, again, at most the first power of the A_2 characteristic of the weight.

Concerning the lower estimate of the square function, it is known that the square function for the standard dyadic filtration on $\mathbb{R}$ satisfies better lower estimates, namely, with a square root on the characteristic instead of linear — the upper and lower estimates estimates are both optimal for the homogeneous filtration, see [5], [12].

Mixed $A_2 - A_\infty$ norm estimates from above were first considered and motivated in [6]. In fact, even the weaker A_∞ characteristic is sufficient for the lower estimate (for the standard dyadic filtration on $\mathbb{R}$), also with square root bounds; this was proved in [17] using the earlier results from [4].

It was a general understanding that in the homogeneous case one should have the same lower bounds as in the case of the standard dyadic filtration on $\mathbb{R}$, but surprisingly, it was not proven before for our "real" square function. The result from [17] gives the desired estimates for a bigger square function, but the statement for our "real" square function (which is the only one that works in the non-homogeneous case) for a homogeneous filtration is proved (to the best of our knowledge) only in the present paper.

The sharp results on the estimates of unconditional basis constants for arbitrary underlying Radon measure and any discrete in time atomic filtration was proved more recently in [13] and then later in [9] by a different method. The constants remain in a linear dependence with the martingale A_2 characteristic, exactly as in the homogeneous situation.

In this paper, we discuss the upper and lower estimates of the square function in this (arbitrary filtration) setting. It is remarkable that the better lower estimates seen in the homogeneous setting fail — indeed the A_∞ bound does not hold true at all — in other words, the A_∞ condition is not sufficient for a lower square function bound. This is even so when using the most restrictive way of defining A_∞. Under the martingale A_2 condition, we obtain a lower estimate, but we will see that it is twice the power of that in the homogeneous case. The failure of the lower estimates motivate us to look closely at the n-adic homogeneous case — one expects a growth with n. Indeed, we show that the lower square function estimate in this setting still holds under the A_∞ assumption, but with a growth $O(n)$.

To see the blow ups we claim, we construct weights, in A_2 or A_∞ respectively, via their martingales based on a filtration where each interval has at most two children, but of possibly very disbalanced measures.

To see the A_∞ lower estimate via the true square function in the n-adic setting, we make use of a Bellman functional taking a distribution function as its variable. This idea stems from [14] — but here is an additional difficulty, similar to that of estimating Haar shifts with Bellman functions.

2. Setup and motivations

2.1. **Filtered atomic spaces.** Let $(\mathcal{X}, \mathcal{F}, \nu)$ be a σ-finite measure space with an atomic filtration, meaning that there exist an increasing sequence of σ-algebras $\mathcal{F}_n$, $n \in \mathbb{N}$ or $n \in \mathbb{Z}$, such that for each n there exists a countable collection $\mathcal{D}_n$ of sets of finite positive measure (called *atoms*) such that $A \in \mathcal{F}_n$ is a union of atoms of $\mathcal{D}_n$.

We will denote $I \in \mathcal{D}_n$ the atoms of $\mathcal{D}_n$, and denote by $\mathcal{D}$ the collection of all atoms, i.e. $\mathcal{D} = \cup_n \mathcal{D}_n$. We allow a set I to belong to several generations $\mathcal{D}_n$, so formally an atom $I \in \mathcal{D}_n$ is a pair (I, n). When there is no confusion, we will omit the "time" n and write simply I instead of (I, n); otherwise when it is necessary to refer to the time n, we will use the symbol $\mathrm{rk}(I)$, such that if I denotes the atom (I, n) then $\mathrm{rk}(I) := n$. Also the inclusion $I \subset J$ for atoms should be understood as inclusion for the sets together with the inequality $\mathrm{rk}(I) \geqslant \mathrm{rk}(J)$. However the union (intersection) of atoms will simply denote the union (intersection) of the corresponding sets regardless of the time component. For $I \in \mathcal{D}_n$ we denote by $\mathrm{ch}(I)$ the set of children of I, that is the atoms of $\mathcal{D}_{n+1}$ that are direct descendants of I : $\mathrm{ch}(I) := \{I' \in \mathcal{D}_{n+1}; I' \subset I\}$.

A typical example is the standard dyadic filtration in $\mathbb{R}^d$ with ν being an arbitrary Radon measure ν; of course, we need to ignore all cubes $Q \in \mathcal{D}$ with $\nu(Q) = 0$.

To avoid nonessential technical details, in this paper we assume that ν is a probability measure, and the filtration is indexed by $n \in \mathbb{Z}_+$. We also assume that $\mathcal{D}_0 = \{\mathcal{X}\}$, and each $\mathcal{D}_n$ is a finite collection (i.e. that every atom has finitely many children).

Since our main results are counterexamples, by providing them in more restrictive settings we get a formally stronger result than in the more general settings. As for the positive estimates, they can be extended to the general case using standard approximation reasoning, so we do not lose anything.

Without loss of generality we can assume that $\mathcal{X}$ is the unit interval $[0, 1]$, the measure ν is the standard Lebesgue measure, and that the atoms are intervals. We assume that the σ-algebra $\mathcal{F}$ is generated by σ-algebras $\mathcal{F}_n$, so more precisely, ν is the restriction of the Lebesgue measure on $\mathcal{F}$.

Measures of intervals are denoted by $|I| := \nu(I)$. For any interval $I \in \mathcal{D}$, we define

$$\langle f \rangle_I = |I|^{-1} \int_I f \mathrm{d}\nu \tag{2.1}$$

and

$$\mathbb{E}_I f = \langle f \rangle_I \mathbf{1}_I.$$

For any interval $I \in \mathcal{D}$, the martingale difference operator Δ_I is defined by

$$\Delta_I f = \sum_{I' \in \mathrm{ch}(I)} \mathbb{E}_{I'} f - \mathbb{E}_I f.$$

Notice that the atom $I \in \mathcal{D}_n$ has only one child (i.e. $\mathrm{ch}(I) = \{I\}$) if and only if the corresponding martingale difference operator is trivial (i.e. $\Delta_I = 0$).

With this in mind, setting

$$\mathbb{E}_n f = \sum_{I \in \mathcal{D}_n} \mathbb{E}_I f = \mathbb{E}(f|\mathcal{F}_n),$$

we define the martingale difference operator Δ_n for any $n > 0$ as

$$\Delta_n f = \mathbb{E}_n f - \mathbb{E}_{n-1} f = \sum_{I\in\mathcal{D}:\mathrm{rk}(I)=n-1} \Delta_I f$$

together with $\Delta_0 f = \mathbb{E}_0 f = \langle f\rangle_{\mathcal{X}}\mathbf{1}$. In the sum above the contributions of the trivial martingale operators is automatically omitted.

For $I \in \mathcal{D}$ denote by D_I the martingale difference space, the image of the operator Δ_I, so $D_I = \Delta_I L^2$ and similarly $D_n = \Delta_n L^2$. Note, that the subspaces D_n, $n \geq 0$ form an orthogonal basis in $L^2 = L^2(\mathcal{X}, \mathcal{F}, \nu)$, and the same holds for the family D_I, $I \in \mathcal{D}$ together with the subspace D_0 (consisting of constants).

2.2. **Bases of martingale difference spaces and the Muckenhoupt A_2 condition.** In the setting described above the following statements are equivalent (with equivalent constants in statements (iii)–(vi)) as a consequence of the general theory of bases, cf. [13].

(i) The system of subspaces $\{D_I : I \in \mathcal{D}, D_I \neq \{0\}\} \cup \{D_0\}$ is an unconditional basis in $L^2(w)$.
(ii) The system of subspaces $\{D_n : 0 \leqslant n < \infty, D_n \neq \{0\}\}$ is an unconditional basis in $L^2(w)$.
(iii) The predictable martingale multipliers T_σ $T_\sigma f = \sum_{I\in\mathcal{D}} \sigma_I \Delta_I f$, with $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$, $\sigma_I \in \{0, 1\}$ (or equivalently $\sigma_I \in \{-1, 1\}$), are uniformly in σ bounded in $L^2(w)$.
(iv) The predictable martingale multipliers T_σ with $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$, $|\sigma_I| \leqslant 1$ are uniformly in σ bounded in $L^2(w)$.
(v) The martingale multipliers T_τ with $\tau = \{\tau_n\}_{n\in\mathbb{N}}$, $\tau_n \in \{0, 1\}$ (or, equivalently $\tau_n \in \{-1, 1\}$),
$$T_\tau f = \sum_{k\in\mathbb{N}} \tau_k \Delta_k f$$
are uniformly in τ bounded in $L^2(w)$.
(vi) The martingale multipliers T_τ with $\tau = \{\tau_n\}_{n\in\mathbb{N}}$, $|\tau_n| \leqslant 1$ are uniformly in τ bounded in $L^2(w)$.

It has been known for some time that the statements (iii)–(vi) hold if and only if the weight w satisfies the martingale Muckenhoupt A_2 condition, see Definition 2.1 below: for the standard dyadic filtration in $\mathbb{R}^N$ we can refer the reader to [7], and for general martingales the result was proved in [2]. Later it was proved that the constants in the statements (iv)–(vi) are estimated by the first power of the A_2 characteristic (i.e. $\lesssim [w]_{2,\mathcal{D}}$): for the standard dyadic filtration in $\mathbb{R}$ (and so in $\mathbb{R}^N$)it was proved in [19]; for the general non-homogeneous filtration it was established in [13] and soon after by a different method in [9].

Let now S denote the square function, as defined in equation (3.1) in Section 3. By taking the average over all $\sigma_I \in \{-1, 1\}$ such as in equation (4.2) one can see that for a weight satisfying the martingale A_2 condition, the quantity $\|Sf\|_{L^2(w)}$ is equivalent in the sense of two sided estimates to the norm $\|f\|_{L^2(w)}$, see the details in Section 4.1.

It can be easily obtained from the estimate $\|T_\sigma\|_{L^2(w)\to L^2(w)} \lesssim [w]_{2,\mathcal{D}}$ that

$$[w]_{2,\mathcal{D}}^{-1}\|f\|_{L^2(w)} \lesssim \|Sf\|_{L^2(w)} \lesssim [w]_{2,\mathcal{D}}\|f\|_{L^2(w)} \qquad \forall f \in L^2(w),$$

see again Section 4.1 for details. The upper bound $\|Sf\|_{L^2(w)} \lesssim [w]_{2,\mathcal{D}}\|f\|_{L^2(w)}$ is known to be sharp, but the lower bound $\|f\|_{L^2(w)} \lesssim [w]_{2,\mathcal{D}}\|Sf\|_{L^2(w)}$, as we discussed above in the introduction, can be improved in the homogeneous case. The investigation of the lower bound in the non-homogeneous situation was the main motivation for this paper.

2.3. **Different A_2 and A_∞ conditions.** Since our underlying filtration can be non-homogeneous, we have to be very careful about the definitions of the classes of weights we will use, as they are no longer necessarily comparable. In all definitions we consider integrable w. Also the notation $\langle\cdot\rangle_I$ below denotes the average operator as defined in (2.1).

Definition 2.1. We say that a weight w satisfies the martingale A_2 condition and write $w \in A_2^{\mathcal{D}}$ if

$$[w]_{2,\mathcal{D}} := \sup_{I\in\mathcal{D}} \langle w\rangle_I \langle w^{-1}\rangle_I < \infty.$$

Definition 2.2. We say that a weight w satisfies the classical A_2 condition and write $w \in A_2^{\mathrm{cl}}$ if

$$[w]_2^{\mathrm{cl}} = \sup_{I\subseteq[0,1]} \langle w\rangle_I \langle w^{-1}\rangle_I < \infty,$$

where the supremum runs over all intervals $I \subset [0,1]$.

Definition 2.3. For an interval I define the *localized* maximal function M_I,

$$M_I f(x) := \mathbf{1}_I(x) \sup_{J\subseteq I:\, x\in J} |\langle f\rangle_J|,$$

where the supremum runs over all intervals $J \subset I$ containing x.

For an interval $I \in \mathcal{D}$ define also the *martingale* localized maximal function $M_I^{\mathcal{D}}$,

$$M_I^{\mathcal{D}} f(x) = \mathbf{1}_I(x) \sup_{J\in\mathcal{D}(I):\, x\in J} \big|\langle f\rangle_J\big|$$

Definition 2.4. We say that a weight w satisfies the classical A_∞ condition and write $w \in A_\infty^{\mathrm{cl}}$ if

$$[w]_{\infty,\mathrm{cl}} = \sup_{I\subseteq[0,1]} \frac{\langle M_I w\rangle_I}{\langle w\rangle_I} < \infty.$$

where $M_I f$ is the localized classical maximal function defined above.

Definition 2.5. We say that a weight w satisfies the *semiclassical* A_∞ condition and write $w \in A_\infty^{\mathrm{scl}}$ if

$$[w]_{\infty,\mathrm{scl}} = \sup_{I\in\mathcal{D}} \frac{\langle M_I w\rangle_I}{\langle w\rangle_I} < \infty,$$

where again $M_I f$ is the classical maximal function localized to $I \in \mathcal{D}$.

Definition 2.6. We say that $w \in A_\infty^{\mathcal{D}}$ if

$$[w]_{\infty,\mathcal{D}} = \sup_{I\in\mathcal{D}} \frac{\langle M_I^{\mathcal{D}} w\rangle_I}{\langle w\rangle_I} < \infty,$$

where $M_I^{\mathcal{D}} f$ is the martingale maximal function localized to $I \in \mathcal{D}$.

We need the following well-known fact.

Proposition 2.7. *For any atomic filtration*

$$[w]_{\infty,\mathcal{D}} \le 4[w]_{2,\mathcal{D}} \tag{2.2}$$

For a simple (but probably not the first) proof see [11, Lemma 4.1]; there it was stated for the standard dyadic filtration on $\mathbb{R}^d$, but the same proof without any changes works for any atomic filtration.

It is a theorem of [13] and [9] that the $A_2^{\mathcal{D}}$ characteristic is sufficient, indeed that the constants above are bounded by a multiple of $[w]_{2,\mathcal{D}}$. It is well known that the $A_2^{\mathcal{D}}$ condition is necessary and that the linear dependence in (2.2) is optimal among all estimates of the form $\Phi([w]_{2,\mathcal{D}})$, which is already seen in the case of dyadic filtration with underlying Lebesgue measure.

3. Main results

For $f \in L^1(\mathcal{X})$ of mean zero the martingale square function is defined by

$$Sf := \left(\sum_I (\Delta_I f)^2\right)^{1/2}. \tag{3.1}$$

For functions that are not of mean zero, the definition is

$$Sf := \left(\mathbb{E}(f)^2 + \sum_I (\Delta_I f)^2\right)^{1/2}. \tag{3.2}$$

For simplicity we consider $\mathcal{X} = [0,1]$ and mean value zero functions. For general functions all our results also hold true if the square function is defined by (3.2).

There are various definitions of the square function in the literature that are not equivalent when the measures are non-homogeneous. Ours is the most natural definition from probability theory, and the only one that works in the non-homogeneous case. For example, for our square the quantity $\|Sf\|_p$ is always equivalent to the norm $\|f\|_p$, $1 < p < \infty$, (with constants depending on p); for other accepted definitions of a square function the equivalence of the norms is true only for homogeneous filtrations, but fails in the non-homogeneous case for $p \neq 2$.

In the paper the expression $A \lesssim B$ means there exists a universal constant c, independent of the important quantities, such as function, weight, measure and filtration, so that $A \leqslant cB$. If the constant depends on some parameters, say a and b, we will write $A \underset{a,b}{\lesssim} B$.

The theorem below is presented just for the sake of completeness. Estimate (3.4) can be easily obtained from known results, see Section 4.1 below. A bit stronger estimate (3.3) can be obtained from the upper bound (Theorem 3.6 below) via Proposition 4.1.

Theorem 3.1. *Given the interval* $[0,1]$ *and any discrete time atomic filtration and any measure, then there holds*

$$\|f\|_{L^2(w)} \lesssim [w]_{2,\mathcal{D}}^{1/2} [w]_{\infty,\mathcal{D}}^{1/2} \|Sf\|_{L^2(w)} \tag{3.3}$$

$$\leq 2[w]_{2,\mathcal{D}} \|Sf\|_{L^2(w)}. \tag{3.4}$$

Here are our main theorems

Theorem 3.2. *The exponent* 1 *of* $[w]_{2,\mathcal{D}}$ *in* (3.4) *is optimal. Namely, given* $A \geq 1$ *one can find a weight* w *defined on the interval* $[0,1]$ *satisfying the classical* A_2 *conditions, such that* $[w]_{2,\mathrm{cl}} = A$ *and a non-homogeneous dyadic filtration* $\mathcal{D}$ *such that for some* $f \in L^2(w)$

$$\|f\|_{L^2(w)} \gtrsim A\|Sf\|_{L^2(w)} = [w]_{2,\mathrm{cl}} \|Sf\|_{L^2(w)};$$

recall that the implied constant here is an absolute one.

Since $[w]_{2,\mathcal{D}} \leqslant [w]_{2,\mathrm{cl}}$ this indeed means that the estimate $\|f\|_{L^2(w)} \lesssim [w]_{2,\mathcal{D}} \|Sf\|_{L^2(w)}$ in Theorem 3.1 is sharp.

Theorem 3.3. *Assumption* $w \in A_\infty^{\mathrm{cl}}$ *is not sufficient for an estimate*

$$\|f\|_{L^2(w)} \leq C([w]_{\infty,\mathrm{cl}}) \|Sf\|_{L^2(w)}.$$

Namely, one can find a weight w *on the interval* $[0,1]$ *satisfying the classical* A_∞ *condition and a non-homogeneous dyadic filtration for which there exists a sequence of functions* $f_n \in L^2(w)$ *with*

$$\|Sf_n\|_{L^2(w)} = 1, \qquad \|f_n\|_{L^2(w)} \to \infty \quad \text{as } n \to \infty.$$

Since $[w]_{\infty,\mathrm{cl}} \geqslant [w]_{\infty,\mathrm{scl}} \geqslant [w]_{\infty,\mathcal{D}}$ this means in particular that no definition of A_∞ is sufficient for a lower square function estimate in the non-homogeneous case.

The following theorem can be obtained combining results from [4] and [17], but here we present a direct proof.

Recall that the n-adic filtration is the atomic filtration where each atom has exactly n children of equal measure.

Theorem 3.4. *For the* n*-adic filtration*

$$\|f\|_{L^2(w)} \lesssim n[w]_{\infty,\mathrm{scl}}^{1/2} \|Sf\|_{L^2(w)}.$$

Remark 3.5. The above theorem holds for an arbitrary *homogeneous* filtration, i.e. for a filtration such that for a certain constant $C_{\mathrm{h}} > 0$,

$$\forall I \in \mathcal{D},\ \forall I' \in \mathrm{ch}(I),\ |I| \leq C_{\mathrm{h}}|I'|.$$

Then it can be seen from the proof that

$$\|f\|_{L^2(w)} \underset{C_{\mathrm{h}}}{\lesssim} [w]_{\infty,\mathrm{scl}}^{1/2} \|Sf\|_{L^2(w)}.$$

In particular, there holds $\|f\|_{L^2(w)} \underset{n}{\lesssim} [w]^{1/2}_{\infty,\mathcal{D}} \|Sf\|_{L^2(w)}$ with additional growth in n.

The following result is probably well-known, see for example [10] for the version for a continuous square function. We present it just for the completeness, and we will just outline the proof of (3.5) in Section 8 and the proof of (3.6) in Section 4.1.

Theorem 3.6. *For an arbitrary atomic filtration and a weight $w \in A_2^{\mathcal{D}}$*

$$\|Sf\|_{L^2(w)} \lesssim [w]^{1/2}_{2,\mathcal{D}} [w^{-1}]^{1/2}_{\infty,\mathcal{D}} \|f\|_{L^2(w)} \tag{3.5}$$

$$\leq 2[w]_{2,\mathcal{D}} \|f\|_{L^2(w)} \tag{3.6}$$

4. Reduction of lower bound to an embedding theorem

It is more convenient to treat the square function S as a linear operator, by paying the price of treating it as an operator to the space of vector-valued functions.

Namely, define $\vec{S} : L^2_0 \to L^2(\ell^2)$ as

$$\vec{S}h = \{\Delta_I h\}_{I\in\mathcal{D}}.$$

Here we treat the sequence $\{\Delta_I h\}_{I\in\mathcal{D}}$ as an element of the ℓ^2-valued space $L^2(\ell^2)$, i.e. we associate with this sequence the function $\vec{S}h$ of two variables, $x \in \Omega$, $k \in \mathbb{N}$,

$$\vec{S}h(x,k) = \Delta_I h(x), \qquad \text{where } I \in \mathcal{D} \text{ is such that } \operatorname{rk}(I) = k.$$

Since for all $x \in \Omega$

$$|Sh(x)| = \|\vec{S}h(x,\,\cdot\,)\|_{\ell^2},$$

we conclude that

$$\|Sh\|_{L^2(w)} = \|\vec{S}h\|_{L^2(w;\,\ell^2)} := \left(\int_\Omega \|\vec{S}h(x,\,\cdot\,)\|^2_{\ell^2} w(x)\mathrm{d}x\right)^{1/2}. \tag{4.1}$$

So the estimates for the square function S are equivalent (with the same constants) to the corresponding estimates for the vector-valued square function $\vec{S}$.

4.1. **Trivial estimates.** Let T_σ, $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$, $\sigma_I \in \{-1,1\}$ be a martingale multiplier,

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

Taking the average $\mathbb{E}_\sigma$ over all possible choices of $\sigma_I \in \{-1,1\}$ (i.e. formally taking σ_I to be independent random variables taking values ± 1 with probability $1/2$), we conclude that for almost all x

$$\mathbb{E}_\sigma \left(|T_\sigma f(x)|^2\right) = \left(Sf(x)\right)^2. \tag{4.2}$$

Therefore, for any weight w and any $f \in L^2(w)$

$$\inf_\sigma \|T_\sigma f\|_{L^2(w)} \leq \|Sf\|_{L^2(w)} \leq \sup_\sigma \|T_\sigma f\|_{L^2(w)}.$$

Thus, denoting by $M(w) := \sup_\sigma \|T_\sigma\|_{L^2(w)\to L^2(w)}$ we can see that

$$M(w)^{-1}\|f\|_{L^2(w)} \le \|Sf\|_{L^2(w)} \le M(w)\|f\|_{L^2(w)}.$$

It is well known that for $w \in A_2^{\mathcal{D}}$

$$\|T_\sigma\|_{L^2(w)\to L^2(w)} \lesssim [w]_{2,\mathcal{D}};$$

for the classical dyadic filtration on $\mathbb{R}$ this result was first proved in [19], and many different proofs are known now for homogeneous filtrations. For the non-homogeneous case it was proved in [13] and then independently and by a different and easier method in [9].

In fact, using the sparse domination technique from [9] one can show that for any atomic filtration one can write the following (stronger) A_2–A_∞ estimate

$$\|T_\sigma\|_{L^2(w)\to L^2(w)} \lesssim [w]_{2,\mathcal{D}}^{1/2}\left([w]_{\infty,\mathcal{D}}^{1/2} + [w^{-1}]_{\infty,\mathcal{D}}^{1/2}\right). \tag{4.3}$$

Another trivial observation is that a lower bound for Sf in $L^2(w)$ can be reduced to the upper bound in $L^2(w^{-1})$:

Proposition 4.1. *Let $w > 0$ a.e. Then*

$$\|f\|_{L^2(w)} \le \|S\|_{L^2(w^{-1})\to L^2(w^{-1})}\|Sf\|_{L^2(w)} \tag{4.4}$$

Proof. By (4.1) estimates for S are reduced to estimating its "linearized" vector-valued version $\vec{S}$. Namely, it is sufficient to estimate the norm in $L^2(w)$ of the *canonical* left inverse $\vec{S}^{-1,\text{left}}$ of $\vec{S}$,

$$\vec{S}^{-1,\text{left}} : \operatorname{Ran}\vec{S} \to L^2;$$

note that since $\vec{S}$ is clearly an injective map, the operator $\vec{S}^{-1,\text{left}}$ is well defined. Note also that there are no weights in the definition of $\vec{S}^{-1,\text{left}}$.

The operator $\vec{S} : L^2 \to L^2(\ell^2)$ (in the non-weighted situation) is an isometry, so

$$\vec{S}^{-1,\text{left}} = \vec{S}^* \Big| \operatorname{Ran}\vec{S}.$$

Therefore

$$\|\vec{S}^{-1,\text{left}}\|_{L^2(w;\,\ell^2)\to L^2(w)} \le \|\vec{S}^*\|_{L^2(w;\,\ell^2)\to L^2(w)}. \tag{4.5}$$

But for an operator $\vec{S}^* : L^2(w;\,\ell^2) \to L^2(w)$ its adjoint with respect to the standard non-weighted duality is the operator $\vec{S} : L^2(w^{-1}) \to L^2(w^{-1};\,\ell^2)$, so

$$\|\vec{S}^{-1,\text{left}}\|_{L^2(w;\,\ell^2)\to L^2(w)} \le \|\vec{S}\|_{L^2(w^{-1})\to L^2(w^{-1};\,\ell^2)},$$

which immediately gives (4.4).

In the above reasoning we skipped a trivial technical detail, namely that $\vec{S}L^2 \ne \vec{S}L^2(w)$ and we have to be a bit careful. However, it all can be fixed by a standard

approximation reasoning. For example, for a finite $\mathcal{F} \subset \mathcal{D}$ we can define the square function $S_{\mathcal{F}}$,

$$S_{\mathcal{F}} h = \left(\sum_{I \in \mathcal{F}} |\Delta_I h|^2 \right)^{1/2}.$$

Then for the vector version $\vec{S}_{\mathcal{F}}$ we do not have a problem with ranges, so the above reasoning gives us the estimate (4.4) with $S_{\mathcal{F}}$ instead of S. Taking the supremum over all finite $\mathcal{F} \subset \mathcal{D}$ we get (4.4). □

4.2. **A sharper way to write the lower bound for the square function.** Analyzing the proof of Proposition 4.1, we can see where one could lose sharpness of the estimate (and in some cases we indeed do lose it): we estimate the norm of the operator $\vec{S}^*$ between weighted spaces, while we need to estimate only the norm of its restriction, which could be smaller.

We wish to find a more convenient equivalent form of the inequality

$$\|h\|_{L^2(w)} \leqslant C \|Sh\|_{L^2(w)} \tag{4.6}$$

that gives us the same constant in the estimate.

Denoting $h_I := \Delta_I h$ the above inequality reads, with the same constant C as above,

$$\left\| \sum_{I \in \tilde{\mathcal{D}}} h_I \right\|_{L^2(w)} = \left\| \sum_{I \in \mathcal{D}} h_I \right\|_{L^2(w)} \leqslant C \left(\sum_{I \in \mathcal{D}} \|h_I\|^2_{L^2(w)} \right)^{1/2} = C \left(\sum_{I \in \tilde{\mathcal{D}}} \|h_I\|^2_{L^2(w)} \right)^{1/2}, \tag{4.7}$$

where we noted in the first and last sum $\tilde{\mathcal{D}} = \{I \in \mathcal{D} : h_I \neq 0\}$.

The standard approximation reasoning implies that it is sufficient to check the above inequality only for finite sums, so we do not have to worry about convergence.

The sequence $\{h_I\}_{I \in \mathcal{D}}$ is a sequence of martingale differences: this simply means that each $h_I = \Delta_I h$ for some h, or, equivalently, that h_I is supported on I, $\int h_I \mathrm{d}x = 0$ and h_I is constant on all $I' \in \mathrm{ch}(I)$.

The above inequality (4.7) holds for all finite sequences $\{h_I\}_{I \in \mathcal{D}}$ of martingale differences if and only the estimate

$$\left\| \sum_{I \in \tilde{\mathcal{D}}} x_I h_I \right\|_{L^2(w)} \leqslant C \left(\sum_{I \in \tilde{\mathcal{D}}} x_I^2 \|h_I\|^2_{L^2(w)} \right)^{1/2} \tag{4.8}$$

holds for all (finite) collections of martingale differences h_I and real numbers x_I, $I \in \tilde{\mathcal{D}}$. The fact that (4.8) implies (4.7) is trivial; on the other hand denoting $x_I h_I$ in (4.8) by h_I we can see that (4.7) implies (4.8).

It looks like we just made the estimate (4.7) more complicated, but this allows us to reduce the problem to a simple "embedding theorem".

Namely, for a fixed sequence $\{h_I\}_{I \in \mathcal{D}}$ of martingale differences let us define the *reconstruction* operator

$$R : \ell^2 = \ell^2(\tilde{\mathcal{D}}) \to L^2, \qquad Rx = \sum_{I \in \tilde{\mathcal{D}}} x_I h_I, \quad \text{where } x = \{x_I\}_{I \in \tilde{\mathcal{D}}}.$$

With respect to the unweighted pairing, its adjoint is the operator

$$R^* : L^2 \to \ell^2, \qquad R^* f = \{(f, h_I)_{L^2}\}_{I\in\tilde{\mathcal{D}}}. \tag{4.9}$$

Define $\gamma = \{\gamma_I\}_{I\in\tilde{\mathcal{D}}} = \left\{\|h_I\|^2_{L^2(w)}\right\}_{I\in\tilde{\mathcal{D}}}$, and the norm in the weighted space $\ell^2(\gamma)$ is given by

$$\|x\|^2_{\ell^2(\gamma)} = \sum_{I\in\tilde{\mathcal{D}}} x_I^2 \gamma_I.$$

The estimate (4.8) can be rewritten as

$$\|Rx\|_{L^2(w)} \le C\|x\|_{\ell^2(\gamma)}.$$

But that is equivalent to the weighted estimate

$$\|R\|_{\ell^2(\gamma)\to L^2(w)} \le C \tag{4.10}$$

For the operator $R : \ell^2(\gamma) \to L^2(w)$ its adjoint with respect to the standard non-weighted duality is the operator

$$R^* : L^2(w^{-1}) \to \ell^2(\gamma^{-1})$$

where $\gamma^{-1} = \{\gamma_I^{-1}\}_{I\in\tilde{\mathcal{D}}}$, and R^* is the adjoint of the operator R in the non-weighted situation ($R : \ell^2 \to L^2$, $R^* : L^2 \to \ell^2$) given by (4.9).

The inequality (4.10) (and so (4.8)) rewritten for the adjoint operator thus becomes

$$\sum_{I\in\tilde{\mathcal{D}}} \frac{(f, h_I)^2_{L^2}}{\gamma_I} \leqslant C^2 \int_0^1 |f|^2 w^{-1}$$

and writing $f = gw$ we can restate it as

$$\sum_{I\in\tilde{\mathcal{D}}} \frac{(g, h_I)^2_{L^2(w)}}{\gamma_I} \leqslant C^2 \int_0^1 |g|^2 w. \tag{4.11}$$

Let us simplify the estimate (4.11) a bit more. Consider the weighted Haar functions h_I^w,

$$h_I^w = h_I - d_I \mathbf{1}_I,$$

where d_I is the unique constant such that $h_I^w \perp \mathbf{1}_I$ in $L^2(w)$. Thanks to orthogonality we have by Pythagorean theorem the estimate $\|h_I^w\|_{L^2(w)} \leqslant \|h_I\|_{L^2(w)}$. Notice further that with this choice of Haar functions, we have $\tilde{\mathcal{D}} = \{I \in \mathcal{D}; \mathrm{ch}(I) \neq I\}$. In particular, if $\mathcal{D}$ is the usual dyadic or n-adic filtration, then $\tilde{\mathcal{D}} = \mathcal{D}$. This is the situation we will consider in the counterexamples built in the next sections.

In order to estimate the sum in (4.11), it suffices to estimate the terms

$$\sum_{I\in\tilde{\mathcal{D}}} \frac{(g, h_I^w)^2_{L^2(w)}}{\gamma_I} \qquad \text{and} \qquad \sum_{I\in\tilde{\mathcal{D}}} \frac{d_I^2 (g, \mathbf{1}_I)^2_{L^2(w)}}{\gamma_I}.$$

The first sum is easily estimated by the Pythagorean theorem:

$$\sum_{I\in\tilde{\mathcal{D}}} \frac{(g,h_I^w)^2_{L^2(w)}}{\gamma_I} = \sum_{I\in\tilde{\mathcal{D}}} \frac{(g,h_I^w)^2_{L^2(w)}}{\|h_I\|^2_{L^2(w)}}$$

$$\leq \sum_{I\in\tilde{\mathcal{D}}} \frac{(g,h_I^w)^2_{L^2(w)}}{\|h_I^w\|^2_{L^2(w)}} \leq \|f\|_{L^2(w)}. \tag{4.12}$$

The second sum can be rewritten as

$$\sum_{I\in\tilde{\mathcal{D}}} \frac{d_I^2\langle gw\rangle_I^2|I|^2}{\gamma_I}$$

and by the martingale Carleson Embedding theorem, it suffices to check its bounds on functions $g=\mathbf{1}_J$, $J\in\tilde{\mathcal{D}}$.

Namely, this sum is bounded by $C_1^2\|f\|^2_{L^2(w)}$ if and only if for all $J\in\tilde{\mathcal{D}}$

$$\frac{1}{|J|}\sum_{I\in\tilde{\mathcal{D}}(J)} \frac{d_I^2\langle w\rangle_I^2|I|^2}{\gamma_I} \leqslant C_2^2\langle w\rangle_J\,. \tag{4.13}$$

Combining this estimate with (4.12) and using the triangle inequality for the ℓ^2 norm, we get that (4.13) holds if and only if

$$\frac{1}{|J|}\sum_{I\in\tilde{\mathcal{D}}(J)} \frac{(w,h_I)^2_{L^2}}{\gamma_I} \leqslant C_3^2\langle w\rangle_J \qquad \forall J\in\tilde{\mathcal{D}}. \tag{4.14}$$

Moreover, we can see that the best constants in inequalities (4.6), (4.13) and (4.14) are equivalent.

5. Counterexample for the A_2 lower bound.

In this section, we will prove Theorem 3.2; note that it is sufficient to prove this theorem for sufficiently large A.

We will first construct a non-homogeneous dyadic filtration on $I_0=[0,1]$ and a weight w with $[w]_{2,\mathrm{cl}}=A$ such that for the best constant C_3 in (4.14) we have for this filtration $C_3\gtrsim[w]_{2,\mathrm{cl}}$. More precisely, we will prove the estimate

$$\sum_{I\in\mathcal{D}(I_0)} \frac{(w,h_I)^2_{L^2}}{\|h_I\|^2_{L^2(w)}} \gtrsim A^2\langle w\rangle_{I_0}\,. \tag{5.1}$$

Then later in Section 5.3 we will show that the weight w we constructed belongs to the classical A_2 class, and that $[w]_{2,\mathrm{cl}}\asymp[w]_{2,\mathcal{D}}$, which completely proves Theorem 3.2.

Note, that since our filtration is dyadic, all martingale difference subspaces $\Delta_I L^2$ are one-dimensional, so the Haar functions h_I are uniquely defined up to a factor. Due to homogeneity of each term in (5.1) a choice of the factor does not matter.

5.1. **Preliminary computations and idea of the proof.** For an interval $I \in \mathcal{D}$ let I_+ and I_- be its children, and let

$$\alpha_\pm^I := |I_\pm|/|I|.$$

The corresponding Haar function h_I is given (up to a constant factor) by

$$h_I = \alpha_-^I \mathbf{1}_{I_+} - \alpha_+^I \mathbf{1}_{I_-}.$$

Then

$$(w, h_I)_{L^2} = \alpha_+^I \alpha_-^I \left(\langle w \rangle_{I_+} - \langle w \rangle_{I_-} \right) |I|,$$

and

$$\|h_I\|_{L^2(w)}^2 = \alpha_+^I \alpha_-^I \left(\alpha_-^I \langle w \rangle_{I_+} + \alpha_+^I \langle w \rangle_{I_-} \right) |I|,$$

so the left hand side in (5.1) is given by

$$\sum_{I \in \mathcal{D}(I_0)} \frac{\alpha_-^I \alpha_+^I \left(\langle w \rangle_{I_+} - \langle w \rangle_{I_-} \right)^2}{\alpha_-^I \langle w \rangle_{I_+} + \alpha_+^I \langle w \rangle_{I_-}} |I|. \tag{5.2}$$

5.1.1. *Idea of the construction.* Assume we have for a term in the sum (5.2) $\alpha_- \ll \alpha_+$ (and in particular $\alpha_-^I \le 0.1$, so $\alpha_+^I \ge 0.9$). Assume also for this term $\alpha_-^I \langle w \rangle_{I_+} \approx \alpha_+^I \langle w \rangle_{I_-}$ so $\langle w \rangle_{I_+} - \langle w \rangle_{I_-} \gtrsim \langle w \rangle_I$, and let also $\langle w \rangle_I |I| \gtrsim \langle w \rangle_{I_0} |I_0|$. Then term we have

$$\frac{\alpha_-^I \alpha_+^I \left(\langle w \rangle_{I_+} - \langle w \rangle_{I_-} \right)^2}{\alpha_-^I \langle w \rangle_{I_+} + \alpha_+^I \langle w \rangle_{I_-}} |I| \gtrsim \langle w \rangle_I |I| \gtrsim \langle w \rangle_{I_0} |I_0|.$$

If we are able to find as many as A^2 such intervals, we will prove (5.1), and therefore also Theorem 3.2.

So let us construct a (non-homogeneous) dyadic filtration $\mathcal{D}$ and a weight $w \in A_2$ such that $[w]_{2,\mathrm{cl}} = A$ such that we have sufficient number of terms as we described above.

In the construction we first show that $[w]_{2,\mathcal{D}} = A$, and later prove that the classical A_2 characteristic remains the same.

5.1.2. *A random walk representation.* To construct a weight we will use its *martingale representation* i.e. get the weight from a random walk in the domain $\Omega_A \subset \mathbb{R}^2$,

$$\Omega_A := \{(u, v) \in \mathbb{R}^2 : 1 \le uv \le A\}.$$

Namely, suppose for each $I \in \mathcal{D}$ we have a point $X_I = (u_I, v_I) \in \Omega_A$, and the points X_I satisfy a (non-homogeneous) *martingale dynamics*,

$$X_I = \alpha_+^I X_{I_+} + \alpha_-^I X_{I_-}; \tag{5.3}$$

here recall $\alpha_\pm^I = |I_\pm|/|I|$.

This collection of points X_I can be interpreted as as a non-homogeneous random walk in Ω_A, where we move from a point X_I to points $X_{I_\pm}$ with probabilities $\alpha_\pm^I$ respectively.

In our example the walk will be stopped after n steps on the lower boundary $uv = 1$ of Ω_A, meaning that for all $I \in \mathrm{ch}^k I_0$, $k > n$ we have

$$u_I v_I = 1.$$

Remark. Note that when the walk hits the lower boundary $uv = 1$ of Ω_A it must stay there; it is immediate corollary of the martingale dynamics (5.3) and the requirement that one must stay above the hyperbola $uv = 1$.

Such a walk immediately gives us a weight $w \in A_2^{\mathcal{D}}$. Namely, take the level N where the walk is stopped on the hyperbola $uv = 1$, and define

$$w := \sum_{I \in \mathrm{ch}^N I_0} u_I \mathbf{1}_I .$$

The martingale dynamics (5.3) together with the fact that $u_I v_I = 1$ for all $I \in \mathrm{ch}^N(I_0)$ imply that for any $I \in \mathcal{D}$

$$\langle w \rangle_I = u_I, \qquad \langle w^{-1} \rangle_I = v_I . \tag{5.4}$$

Since $X_I \in \Omega_A$, identities (5.4) mean that $[w]_{2,\mathcal{D}} \le A$; if we, for example start the walk at a point on the upper hyperbola $uv = A$, then trivially $[w]_{2,\mathcal{D}} = A$.

5.2. **The construction.** Let us construct the non-homogeneous dyadic filtration and the corresponding random walk in Ω_A, which gives us the weight w as follows.

5.2.1. *Setting up the random walk.* We restrict our attention to the one dimensional dyadic setting. Let $I_0 = [0, 1]$. The dyadic filtration $\mathcal{D}(I_0)$ is such that each $I \in \mathcal{D}$ has exactly 2 children, I_+ and I_-, with equal Lebesgue measure $\lambda(I_-) = \lambda(I_+) = \lambda(I)/2$. However, with respect to the non homogeneous measure ν, we have $\nu(I\pm) := |I_\pm| := \alpha_\pm^I |I|$, and we will be choosing the probabilities $\alpha_\pm^I$ in order to completely define the dyadic lattice.

For easier bookkeeping let I_+ always be on the right, and let $|I_+| \ge |I_-|$.

We start from the interval $I_0 = [0, 1]$, and pick a point $X_0 = X_{I_0} = (u_0, v_0)$ on the upper hyperbola $uv = Q_0 = A$. We will then construct the random walk in such a way, that at each interval I anything interesting can happen only on its right part I_+; on the left part I_- the walk stops on the lower hyperbola $uv = 1$. Because we are stopped on the lower hyperbola, it does not matter how we continue the filtration $\mathcal{D}$ on I_-; we can, for example continue it as the standard dyadic filtration.

So, we start from the interval I_0, and anything interesting will happen only on its right part $(I_0)_+ =: I_1$, because the walk will stop on $(I_0)_- =: I_1^\star$. We then split the interesting interval I_1 into two parts $I_2 := (I_1)_+$ and $I_2^\star := (I_1)_-$, so again on $I_2^\star$ the walk stops, and so on...

So, we will only need to keep track of what is going on on intervals I_k, $I_k^\star$, $k \ge 1$

$$I_{k+1} := (I_k)_+, \quad I_{k+1}^\star := (I_k)_-, \qquad k \ge 0.$$

Denoting for simplification of notation the corresponding probabilities $\alpha_\pm^I$ by α_k and $\alpha_k^\star$, we write

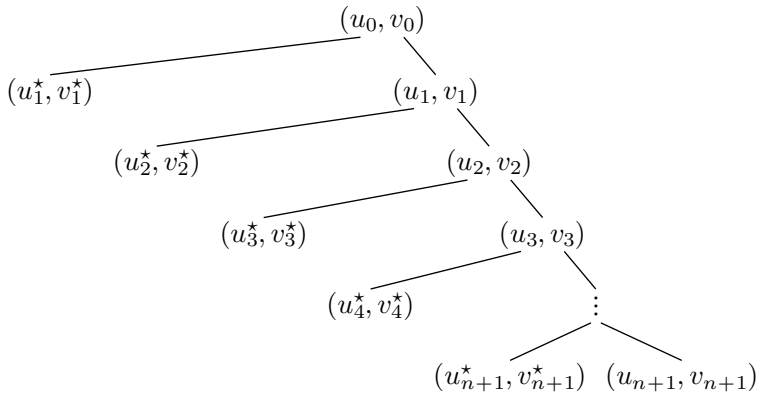
$$|I_{k+1}| = \alpha_k |I_k|, \quad |I_{k+1}^\star| = \alpha_k^\star |I_k|, \qquad k \ge 0$$

(clearly $\alpha_k + \alpha_k^* = 1$); the values of α_k, $\alpha_k^\star$ will be chosen later.

The points $X_k = (u_k, v_k)$, $X_k^\star = (u_k^\star, v_k^\star)$ of our walk must satisfy the martingale dynamics (5.3), which in our notation can be rewritten as

$$X_k = \alpha_k X_{k+1} + \alpha_k^\star X_{k+1}^\star. \tag{5.5}$$

Schematically, the random walk we need to track can be presented in the picture below.



5.2.2. *Inductive construction.* We start from a point $X_0 = (u_0, v_0)$, $u_0 v_0 = Q_0 := A$, and construct the the walk by induction. Suppose we constructed the points $X_1, X_2, \dots, X_k$, and $X_1^\star, X_2^\star, \dots, X_k^\star$, and let $Q_k := u_k v_k$. We will continue our iterations as long as $Q_k \ge Q_0/2$; if $Q_k < Q_0/2$ we stop the walk by moving from the point X_k to the both points being on the lower hyperbola $uv = 1$.

If $Q_k \ge Q_0/2$ we set

$$\alpha_k^\star = 1/Q_k, \qquad \alpha_k = 1 - \alpha_k^\star. \tag{5.6}$$

The point $X_{k+1}^\star$ is defined as the point of intersection of the tangent line to the hyperbola $uv = Q_k$ at the point $X_k = (u_k, v_k)$ and the lower hyperbola $uv = 1$. The computations show

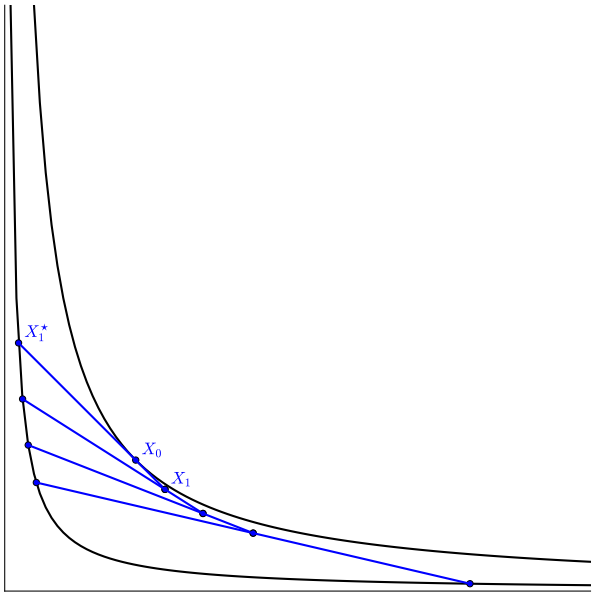
$$u_{k+1}^\star = \left(1 - \sqrt{1 - 1/Q_k}\right) u_k, \qquad v_{k+1}^\star = \left(1 + \sqrt{1 - 1/Q_k}\right) v_k;$$

probably the easiest way to compute is to do first the computations for the case $u_k = v_k = Q_k^{1/2}$ and then do the rescaling $u \mapsto \lambda u$, $v \mapsto \lambda^{-1} v$ for an appropriate λ.

It follows from the martingale dynamics (5.5) that

$$\begin{aligned} u_{k+1} &= \left(1 + \frac{\alpha_k^\star}{\alpha_k}\sqrt{1 - 1/Q_k}\right) u_k, \\ &= \left(1 + \alpha_k^\star \alpha_k{}^{-1/2}\right) u_k, \end{aligned} \qquad \begin{aligned} v_{k+1} &= \left(1 - \frac{\alpha_k^\star}{\alpha_k}\sqrt{1 - 1/Q_k}\right) v_k, \\ &= \left(1 - \alpha_k^\star \alpha_k{}^{-1/2}\right) v_k. \end{aligned}$$

The figure below shows an example of a dyadic martingale as above with $X_k = (u_k, v_k)$ with $0 \leqslant k \leqslant 4$, $X_k^\star = (u_k^\star, v_k^\star)$, with $1 \leqslant k \leqslant 3$. Only X_0, X_1 and $X_0^\star$ are labelled. The two hyperbolas are $uv = 1$ and $uv = Q_0 = A$. All the points lie in the domain Ω_A.



5.2.3. *The estimates.* Let us now write some estimates. Let us assume that $Q_0 = A \geq 4$, so $Q_k \geq A/2 = Q_0/2 \geq 2$. Then

$$\begin{aligned} u_{k+1} - u_{k+1}^\star \geq u_k - u_{k+1}^\star = u_k\sqrt{1-1/Q_k} \geq u_k/\sqrt{2}, \\ \alpha_k^\star u_{k+1} + \alpha_k u_{k+1}^\star = \left[\alpha_k^\star(1+\alpha_k^\star\alpha_k^{-1/2}) + \alpha_k(1-\alpha_k^{1/2})\right] u_k \\ \leq \left[\alpha_k^\star(1+\alpha_k^\star\alpha_k^{-1/2}) + \alpha_k^\star\alpha_k\right] u_k \lesssim \alpha_k^\star u_k. \end{aligned}$$

Combining the above estimates together we get that

$$\frac{\alpha_k\alpha_k^\star(u_{k+1}-u_{k+1}^\star)^2}{\alpha_k^\star u_{k+1} + \alpha_k u_{k+1}^\star}|I_k| \gtrsim u_k|I_k| \tag{5.7}$$

Using formulas for u_{k+1} and $u_{k+1}^\star$ we get that

$$\begin{aligned} Q_{k+1} &= \left(1+\alpha_k^\star{\alpha_k}^{-1/2}\right)\left(1-\alpha_k^\star{\alpha_k}^{-1/2}\right) Q_k \\ &= \left(1 - Q_k^{-2}(1-1/Q_k)^{-1}\right) Q_k \\ &\geq \left(1-2Q_k^{-2}\right) Q_k \geq \left(1-8Q_0^{-2}\right) Q_k\,. \end{aligned} \tag{5.8}$$

Finally, since $u_{k+1} = (1+\alpha_k^\star\alpha_k^{-1/2})u_k$ we get

$$\begin{aligned} u_{k+1}|I_{k+1}| &= (1-\alpha_k^\star)(1+\alpha_k^\star\alpha_k^{-1/2})u_k|I_k| \\ &\geq \left(1-(\alpha_k^\star)^2\right) u_k|I_k| = \left(1-1/Q_k^2\right) u_k|I_k| \\ &\geq \left(1-4/Q_0^2\right) u_k|I_k|. \end{aligned} \tag{5.9}$$

The estimate (5.8) implies that

$$Q_k \geq \left(1 - 8Q_0^{-2}\right)^k Q_0,$$

so for $n \gtrsim Q_0^2$ steps we will have $Q_k \geq Q_0/2$, $k \leq n$. Finally, it follows from (5.9) that

$$u_k|I_k| \geq \left(1 - 4/Q_0^2\right)^k u_0|I_0|,$$

therefore $u_k|I_k| \geq \frac{1}{2}u_0|I_0|$ for $k \leq n$. From (5.7) we get that for $k \leq n$

$$\frac{\alpha_k\alpha_k^\star(u_{k+1} - u_{k+1}^\star)^2}{\alpha_k^\star u_{k+1} + \alpha_k u_{k+1}^\star}|I_k| \gtrsim u_0|I_0|$$

5.2.4. *Finishing the random walk.* First of all let us note that in our construction not only the points X_k, $X_k^\star$, but the whole interval $[X_k, X_k^\star]$ are in the domain Ω_{A}. That will be needed in proving that the weight w we constructed satisfies the classical A_2 condition and that $[w]_{2,\mathcal{D}} = [w]_{2,\mathrm{cl}}$.

Note also that the following follows immediately from the construction:

(i) The sequence u_k is increasing, the sequence v_k is decreasing.
(ii) The sequence Q_k is decreasing.
(iii) The slopes of intervals $[X_k^\star, X_k]$ are negative and increasing (i.e. have decreasing absolute values).

In our construction we made n steps while $Q_k \geq Q_0/2$. Now we need to stop the process by moving from X_n to the points X_{n+1}, $X_{n+1}^\star$ on the lower hyperbola $uv = 1$. Note that we can easily do it preserving the above properties (i)–(iii); recall that we have a choice of transition probabilities α_n, $\alpha_n^\star$.

5.3. **Why the constructed weight belongs to classical A_2.** It is of independent interest to observe that even classical A_2^{cl}, containing many more intervals as competitors, is not sufficient for a square root bound. We will show that the example above indeed belongs to the classical A_2 and that $[w]_{2,\mathcal{D}} = [w]_{2,\mathrm{cl}}$.

The following argument is borrowed from [8]. Let $X : I_0 \to \mathbb{R}^2$ be a vector-valued function, $X(t) = (w(t), w(t)^{-1})$.

Consider the trajectory

$$\gamma(t) := \langle X\rangle_{[t,1]}, \qquad t \in I_0 = [0,1].$$

Notice that $\gamma(0) = (w_0, v_0)$ is the starting point. Let β_k be the left endpoint of the interval I_k, then

$$\gamma(\beta_k) = (1 - \beta_k)X_k, \qquad X_k = (u_k, v_k). \tag{5.10}$$

Since the weight is constant on the interval $I_{k+1} \setminus I_k$ we see that on this interval the trajectory of $\gamma(t)$ in the uv plane is exactly the line segment joining the points X_k and X_{k+1} (note that this segment is the part of the interval $[X_k^\star, X_k]$).

Indeed, since both w and w^{-1} are constant on $I_{k+1} \setminus I_k$, both u and v coordinates of $\gamma(t)$ have a form

$$\frac{a + bt}{1 - t} = \frac{a + b}{1 - t} - b,$$

so both coordinates are affine functions of the variable $s = 1/(1-t)$. Therefore the trajectory indeed lies on a line segment. The monotonicity of the change of variables $s = 1/(1-t)$ together with (5.10) insure that this segment is exactly $[X_k, X_{k+1}]$.

Clearly the trajectory of $\gamma(t)$ is convex (increasing slopes, see (iii) in Section 5.2.4 above), piecewise linear, and it belongs to the domain

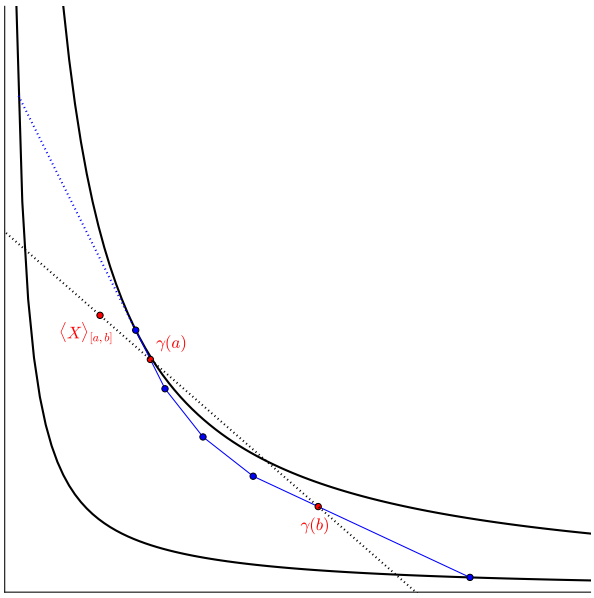
$$\Omega_A := \{(u,v) \in \mathbb{R}^2 : 1 \le uv \le A\}.$$

The line segments at the endpoints of the curve γ if extended to the line liees below the graph $uv = A$ (here we should agree that on the final interval I_n we concatenated the weight along the line segment not intersecting the previous line segments and the boundary $uv = A$).

Take arbitrary $1 \ge b > a \ge 0$. Since

$$\gamma(a) = \frac{1-b}{1-a} \cdot \gamma(b) + \frac{b-a}{1-a} \cdot \langle X \rangle_{[a,b]},$$

it follows from a simple geometry that $\langle X \rangle_{[a,b]} \in \Omega_{Q_0}$. The figure below illustrates the equation above. Notice that the segment $[\langle X \rangle_{[a,b]}, \gamma(a)]$ lies below the convex curve $\gamma(t)$ and below its tangent at $t = 0$. This ensures that $\langle X \rangle_{[a,b]}$ belongs to Ω_A.



6. No bounds in terms of A_∞

In this section we prove Theorem 3.3. We show that in the non-homogeneous setting, if $[w]_{\infty,\mathrm{cl}} < \infty$ then we can choose a filtration so that the sum

$$\frac{1}{|J|}\sum_{I\subseteq J}\frac{\alpha_+^I\alpha_-^I(\langle w\rangle_{I_+}-\langle w\rangle_{I_-})^2}{\alpha_+^I\langle w\rangle_{I_-}+\alpha_-^I\langle w\rangle_{I_+}}|I|$$

can be very large (so no bound in terms of $A_{\infty,\mathrm{cl}}$ characteristics can be obtained).

Indeed, Take $w(x)=x$ on $[0,1]$. It is not difficult to check that $[w]_{\infty,\mathrm{cl}}$ is finite. Let $\varepsilon>0$ be a sufficiently small number (we will specify it later). We will construct the filtratrion as follows (parent $\to$ children)

$$\begin{aligned}
&I_0:=[0,1]; && I_0^-:=[0,\varepsilon], && I_0^+:=[\varepsilon,1];\\
&I_1:=I_0^+; && I_1^-:=[\varepsilon,2\varepsilon], && I_1^+:=[2\varepsilon,1];\\
&\dots\\
&I_{k-1}:=[(k-1)\varepsilon,1]; && I_{k-1}^-:=[(k-1)\varepsilon,k\varepsilon], && I_{k-1}^+:=[k\varepsilon,1]
\end{aligned}$$

Then

$$\begin{aligned}
\langle w\rangle_{I_{k-1}^-}&=\frac{\varepsilon(2k-1)}{2}; & \langle w\rangle_{I_{k-1}^+}&=\frac{1+\varepsilon k}{2};\\
\alpha_{I_{k-1}^-}&:=\frac{\varepsilon}{1-\varepsilon(k-1)}; & \alpha_{I_{k-1}^+}&:=\frac{1-\varepsilon k}{1-\varepsilon(k-1)}.
\end{aligned}$$

Let's say we make N steps. Then

$$\sum_{k=1}^N\frac{\alpha_{I_{k-1}^-}\alpha_{I_{k-1}^+}(\langle w\rangle_{I_{k-1}^-}-\langle w\rangle_{I_{k-1}^+})^2}{\alpha_{I_{k-1}^-}\langle w\rangle_{I_{k-1}^+}+\alpha_{I_{k-1}^+}\langle w\rangle_{I_{k-1}^-}}|I_{k-1}|=\frac12\sum_{k=1}^N\frac{(1-\varepsilon k)(1-\varepsilon(k-1))^2}{(1+\varepsilon k)+(1-\varepsilon k)(2k-1)}.$$

Choose $\varepsilon=\frac{1}{N}$. Then

$$\begin{aligned}
\frac12\sum_{k=1}^N\frac{(1-\varepsilon k)(1-\varepsilon(k-1))^2}{(1+\varepsilon k)+(1-\varepsilon k)(2k-1)} &\geqslant \frac18\sum_{k=1}^N\frac{(1-k/N)^3}{k}\\
&\geqslant \frac18\sum_{k=1}^N\frac{1-3k/N}{k}\\
&\geqslant \frac18(\ln(N-1)-3)
\end{aligned}$$

and it becomes very large as $N\to\infty$.

7. Estimate in terms of martingale $A_\infty^{\mathcal{D}}$ for homogeneous filtrations

In this section we prove Theorem 3.4.

Since everything scales correctly, we can assume without loss of generality that the starting interval I_0 of our filtration is $I_0=[0,1]$.

Let $\mathcal{D} = \mathcal{D}(I_0)$ denote all n-adic intervals $I \subset I_0$.

7.1. **Bellman functional and its properties.** For a non-negative function w on an interval I let $N = N_I^w$ be its *normalized* distribution function,

$$N_I^w(t) := |I|^{-1} \left|\{x \in I : w(x) > t\}\right|, \qquad t \geq 0, \tag{7.1}$$

Trivially the normalized distribution function N_I^w satisfies the *martingale dynamics*, namely, if I_k are the children of I, then

$$N_I^w = \sum_k \alpha_k N_{I_k}^w, \qquad \text{where } \alpha_k = |I_k|/|I|.$$

On the set of distribution functions consider the Bellman functional

$$B(N) = \int_0^\infty \psi(N(t))\mathrm{d}t$$

with $\psi(s) = s - s\ln(s)$.

We will need the following well-known fact, see [1, Theorem IV.6.7].

Lemma 7.1. *Let w be a non-negative function on $I_0 = [0,1]$ and let $N = N_{I_0}^w$ be its distribution function. Then $\|M_{I_0} w\|_{L^1}$ and $B(N)$ are equivalent in the sense of two-sided estimates (with some absolute constants).*

Let $N = N_0$ and N_1 be two distribution functions, and let $\triangle N := N_1 - N$. We want to compute the second derivative of the function $\theta \mapsto B(N + \theta \triangle N)$.

Let $N_\theta := N + \theta \triangle N$, and let

$$u_\theta := \int_0^\infty N_\theta(t)dt.$$

If we think of the function N_θ as of the distribution function of a function w_θ on, say, $[0,1]$, then u_θ is the average of the function w_θ. Also, denote

$$\triangle u := u_1 - u_0 = \int_0^\infty \triangle N(t)\mathrm{d}t. \tag{7.2}$$

Then we calculate

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} B(N_\theta) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \int_0^\infty \psi(N_\theta(t))\mathrm{d}t = -\int_0^\infty \frac{(\triangle N(t))^2}{N_\theta(t)}\mathrm{d}t.$$

Using the Cauchy–Schwartz inequality we get, see [14, Lemma 5.1], that

$$-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} B(N_\theta) \geqslant \frac{\left(\int_0^\infty \triangle N(t)\mathrm{d}t\right)^2}{\int_0^\infty N_\theta(t)\mathrm{d}t} = \frac{|\triangle u|^2}{u_\theta}.$$

Then using the Taylor's formula we get, see [14, Corollary 5.2]

Lemma 7.2. *Let N_1, N_2 and N be distribution functions such that $N = (N_1 + N_2)/2$ and $N = N(N_{1,2}) < \infty$. Let $\triangle N = N_1 - N$ and $\triangle u$ is defined by (7.2). Then*

$$B(N) - \frac{B(N_1) + B(N_2)}{2} \geq \frac{1}{2} \cdot \frac{(\triangle u)^2}{u}, \tag{7.3}$$

where, recall $u = \int_0^\infty N(t)\mathrm{d}t$.

Using this lemma one can easily get the result for the dyadic filtration. To get it for the n-adic filtration some extra work is needed.

Definition 7.3. Recall that a Haar function on an interval $I \in \mathcal{D}$ is a function $h = h_I$ supported on I, constant on children of I and such that $\int_I h_I \mathrm{d}x = 0$.

A Haar function h_I is called elementary if it is non-zero on at most 2 children of I. Thus any elementary Haar function h_I can be represented as $h_I = c_I \left(\mathbf{1}_{I_{k_1}} - \mathbf{1}_{I_{k_2}}\right)$, $I_{k_1}, I_{k_2} \in \operatorname{ch} I$.

Lemma 7.4. *Let $\mathcal{D}$ be an n adic filtration. Any Haar function h on an interval $I \in \mathcal{D}$ can be represented as a sum of at most n elementary Haar functions h_k, and moreover*

$$|h| = \sum_k |h_k| \tag{7.4}$$

Proof. We prove it using induction in n. The case $n = 2$ is trivial.

Suppose the lemma is proved for $n-1$. Let I_k be the children of I. We write h as

$$h = \sum_{k=1}^{n} \eta_k \mathbf{1}_{I_k}.$$

Since $\int_h \mathrm{d}x = 0$ there exist k_1, k_2 such that $\eta_{k_1} > 0$, $\eta_{k_2} < 0$.

For $\mu_1 := \min(|\eta_{k_1}|, |\eta_{k_2}|)$ define

$$h_1 := \mu_1 \left(\mathbf{1}_{I_{k_1}} - \mathbf{1}_{I_{k_2}}\right), \qquad h^1 := h - h_1.$$

Clearly, h_1 is an elementary Haar function, h is a Haar function and

$$|h| = |h_1| + |h^1|. \tag{7.5}$$

Note, that h^1 is supported on at most $n-1$ intervals. Applying the induction hypothesis we get the decomposition $h = \sum_k h_k$. Identity (7.4) follows from (7.5). □

7.2. **Proof of Theorem 3.4.** We need to estimate the left hand side of (4.14), i.e. the sum

$$\frac{1}{|I_0|} \sum_{I \in \mathcal{D}(I_0)} \frac{(w, h_I)^2_{L^2}}{\|h_I\|^2_{L^2(w)}} \tag{7.6}$$

Recall that for an interval $I \in \mathcal{D}_I$, we note N_I^w the distribution function (7.1). We want to show that

$$|I| B(N_I) - \sum_{I_k \in \operatorname{ch} I} |I_k| B(N_{I_k}) \geq \frac{2}{n^2} \frac{(w, h_I)^2_{L^2}}{\|h_I\|^2_{L^2(w)}} \tag{7.7}$$

Then summing over all $I \in \mathcal{D}(I_0)$ and taking into account that $B(N_I) \geq 0$ we get that

$$\sum_{I \in \mathcal{D}(I_0)} \frac{(w, h_I)^2_{L^2}}{\|h_I\|^2_{L^2(w)}} \leq \frac{n^2}{2} B(N_{I_0}) \lesssim n^2 \|M_{I_0} w\|_{L^1(I_0)};$$

the last inequality here follows from Lemma 7.1. By the definition of A_∞

$$\|M_{I_0} w\|_{L^1(I_0)} \le [w]_{\infty,\mathrm{cl}} \langle w\rangle_{I_0} |I_0| = [w]_{\infty,\mathrm{cl}} \langle w\rangle_{I_0},$$

so the theorem is proved modulo the main inequality (7.7).

To proof (7.7) let us decompose the Haar function h_I into the sum of elementary Haar functions $h_{I,k}$, $h = \sum_k h_{I,k}$, see Lemma 7.4.

It follows from (7.4) that

$$\|h_{I,k}\|_{L^2(w)} \le \|h_I\|_{L^2(w)}. \tag{7.8}$$

Certainly

$$(w, h_I)_{L^2} = \sum_{k=1}^{n} (w, h_{I,k})_{L^2},$$

so there exists a k so that

$$|(w, h_{I,k})_{L^2}| \geqslant \frac{1}{n} |(w, h_I)_{L^2}|. \tag{7.9}$$

Without loss of generality (by rearranging the intervals, if necessary) we can assume that this $k = 1$ and that the elementary Haar function $h_{I,1}$ is a dyadic Haar function supported on the first two n-adic subintervals I_1 and I_2 of I.

Denote $I^1 = I_1 \cup I_2$. Then

$$N_I = \frac{2}{n} N_{I^1} + \frac{1}{n} \sum_{k=1}^{n} N_{I_k}, \qquad \text{and} \qquad N_{I^1} = \frac{1}{2}\left(N_{I_1} + N_{I_2}\right)$$

By concavity of B we get

$$|I| B(N_I) \geqslant \sum_{k=3}^{n} \frac{|I|}{n} B(N_k) + \frac{2}{n} |I| B\left(\frac{N_1 + N_2}{2}\right).$$

Note that for the elementary Haar function $h_{I,1}$

$$\frac{(w, h_{I,1})^2_{L^2}}{\|h_{I,1}\|^2_{L^2(w)}} = \frac{\left(\langle w\rangle_{I_1} - \langle w\rangle_{I_2}\right)^2}{\langle w\rangle_{I^1}} |I^1| = 4 \frac{\left(\langle w\rangle_{I_1} - \langle w\rangle_{I^1}\right)^2}{\langle w\rangle_{I^1}} |I^1|$$

Then applying Lemma 7.2 and noticing that Δu in (7.3) us exactly $\langle w\rangle_{I_1} - \langle w\rangle_{I^1}$ we get

$$\begin{aligned} |I^1| B\left(\frac{N_1 + N_2}{2}\right) &\geqslant \frac{|I^1|}{2}(B(N_1) + B(N_2)) + 2 \frac{(w, h_{I,1})^2_{L^2}}{\|h_{I,1}\|^2_{L^2(w)}} && \text{by (7.3)} \\ &\geqslant \frac{|I^1|}{2}(B(N_1) + B(N_2)) + \frac{2}{n^2} \frac{(w, h_I)^2_{L^2}}{\|h_I\|^2_{L^2(w)}} && \text{by (7.8) and (7.9).} \end{aligned}$$

The main inequality (7.7), and so the theorem is proved.

7.3. **Some remarks.** It is a remarkable result of [4] that for any $Q \subset \mathbb{R}^n$ we have superexponential bound

$$\frac{1}{|Q|}\,|\{x \in Q \ : \ f(x) - \langle f\rangle_Q \geq \lambda\}| \leq e^{-\lambda^2/(2\|S_\infty f\|_\infty^2)} \tag{7.10}$$

for any $\lambda \geq 0$ and any f with $\|S_\infty f\|_\infty < \infty$, where the square function S_∞ is defined as follows

$$S_\infty f = \left(\sum_{I\in\mathcal{D}(Q)} \|\Delta_I f\|_\infty^2 \mathbf{1}_I\right)^{1/2}.$$

The superexponential estimate allowed Wilson [18] to obtain weighted L^p estimates for the square function in terms of the maximal function, namely for any $0 < p < \infty$ we have

$$\int |M_{\mathcal{D}} f|^p w\mathrm{d}x \underset{n,p}{\lesssim} [w]_\infty^{p/2} \int (S_\infty f)^p w\mathrm{d}x \tag{7.11}$$

For the standard dyadic filtration S_∞ coincides with our square function S, so the result of Wilson (for $p = 2$) gives for the standard dyadic filtration the statement of Theorem 3.4. However, this approach does not give Theorem 3.4 for n-adic filtration with $n \geq 3$, because the superexponential estimate (7.10) should be first proved for our square function S. And the square function S_∞ is significantly larger than S: one can easily construct an example of a function with $\|Sf\|_\infty \leq 1$ and unbounded $S_\infty f$. So Theorem 3.4 is a new result.

We should mention that it is possible using some ideas from the proof of Theorem 3.4 to prove the estimate (7.10) for our square function S. The reasoning from [18] then allows us to get the estimate (7.11) for our square function, but this will be a subject of a separate paper.

8. Upper bound for the square function

In this section we sketch a proof of the harder estimate (3.5) in Theorem 3.6; the easier estimate (3.6) was proved earlier in Section 4.1.

Trivial reasoning shows that it is sufficient to prove the estimate for an atomic filtration on $I_0 = [0, 1]$.

The proof is based on the sparse domination of the square function.

Recall that a collection $\mathcal{S} \subset \mathcal{D}$ is called *sparse* if for any $J \in \mathcal{S}$

$$\sum_{I\in \mathrm{ch}_{\mathcal{S}} J} |I| \leq |J|/2.$$

Given a sparse family $\mathcal{S}$ the *sparse square function* $S_{\mathcal{S}}$ is defined as

$$S_{\mathcal{S}} f(x) := \left(\sum_{I\in\mathcal{S}} \langle |f|\rangle_I^2 \mathbf{1}_I(x)\right)^{1/2}$$

Lemma 8.1. *Let $f \in L^1(I_0)$. There exist a sparse collection $\mathcal{S} \subset \mathcal{D}$ (depending on f) such that*

$$Sf(x) \lesssim S_{\mathcal{S}} f(x) \qquad a.e.$$

Proof. The construction is pretty standard, we just outline it.

It is well known that the operator S has weak type 1-1, see [3]. The maximal function $M^{\mathcal{D}}$ also has weak type 1-1, so there exists constant C such that

$$\Big|\{x \in J : S_J f(x) > C\} \bigcup \{x \in J : M^{\mathcal{D}}_J f(x) > C\}\Big| \le |J|/2; \tag{8.1}$$

here S_J is the *localized* square function

$$S_J f(x) := \Big(\sum_{I \in \mathcal{D}(J)} |\Delta_I f(x)|^2\Big)^{1/2}.$$

We start from the interval I_0. We define the stopping intervals $I \in \mathcal{S}_1(I_0)$ to be the maximal (by inclusion) intervals $I \in \mathcal{D}(I_0)$ such that either

$$\langle |f| \rangle_I > C \langle |f| \rangle_{I_0} \qquad \text{or} \qquad \sum_{J \in \mathcal{D}(I_0): I \subsetneq J} |\Delta_J f(x)|^2 > C^2 \langle |f| \rangle^2_{I_0};$$

here C is from (8.1) and clearly $S = S_{I_0}$.

By (8.1) we have $\sum_{I \in \mathcal{S}_1(I_0)} |I| \le |I_0|/2$, and

$$Sf(x)^2 \le 3C^2 \langle |f| \rangle^2_{I_0} \mathbf{1}_{I_0} + 2C^2 \sum_{I \in \mathcal{S}_1(I_0)} \langle |f| \rangle_I \mathbf{1}_I + \sum_{I \in \mathcal{S}_1(I_0)} S_I f(x)^2.$$

Repeating this procedure for stopping intervals $I \in \mathcal{S}_1(I_0)$ and iterating, we get the conclusion of the lemma. □

Proof of estimate (3.5). It is sufficient to show that for a sparse family $\mathcal{S}$

$$\|S_{\mathcal{S}} f\|_{L^2(w)} \lesssim [w]^{1/2}_{2,\mathcal{D}} [w^{-1}]^{1/2}_{\infty,\mathcal{D}} \|f\|_{L^2(w)}$$

Denoting $g = wf$, so $f = w^{-1}g$ we can rewrite this estimate as

$$\|S_{\mathcal{S}}(gw^{-1})\|_{L^2(w)} \lesssim [w]^{1/2}_{2,\mathcal{D}} [w^{-1}]^{1/2}_{\infty,\mathcal{D}} \|g\|_{L^2(w^{-1})} \tag{8.2}$$

So, we need to estimate

$$\sum_{I \in \mathcal{S}} \langle |g| w^{-1} \rangle^2_I \langle w \rangle_I |I| \tag{8.3}$$

(the left hand side in (8.2) squared). But as we already discussed above in Section 4.2, the martingale Carleson Embedding Theorem implies that it is sufficient to estimate (8.3) on functions $g = \mathbf{1}_J$, $J \in \mathcal{D}$. Namely, if for all $J \in \mathcal{D}$

$$\sum_{I \in \mathcal{S}: I \subset J} \langle w^{-1} \rangle^2_I \langle w \rangle_I |I| \le C \langle w^{-1} \rangle_J |J|$$

then for all $g \in L^2(w^{-1})$, the sum (8.3) is bounded by $4C\|g\|^2_{L^2(w^{-1})}$.

Estimating we get

$$\begin{aligned}\sum_{I\in\mathcal{S}:I\subset J}\langle w^{-1}\rangle_I^2\langle w\rangle_I|I| &\le [w]_{2,\mathcal{D}}\sum_{I\in\mathcal{S}:I\subset J}\langle w^{-1}\rangle_I|I|\\ &\le [w]_{2,\mathcal{D}}\|M_J(w^{-1})\|_{L^1}\\ &\le [w]_{2,\mathcal{D}}[w^{-1}]_{\infty,\mathcal{D}}.\end{aligned}$$

□

References

[1] C. Bennett and R. Sharpley, *Interpolation of operators*, Pure and Applied Mathematics, vol. 129, Academic Press Inc., Boston, MA, 1988.

[2] A. Bonami and D. Lépingle, *Fonction maximale et variation quadratique des martingales en présence d'un poids*, Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78), Lecture Notes in Math., vol. 721, Springer, Berlin, 1979, pp. 294–306.

[3] D. L. Burkholder, *Martingale transforms*, Ann. Math. Statist. **37** (1966), 1494–1504.

[4] S.-Y. A. Chang, J. M. Wilson, and T. H. Wolff, *Some weighted norm inequalities concerning the Schrödinger operators*, Comment. Math. Helv. **60** (1985), no. 2, 217–246.

[5] S. Hukovic, S. Treil, and A. Volberg, *The Bellman functions and sharp weighted inequalities for square functions*, Complex analysis, operators, and related topics, Oper. Theory Adv. Appl., vol. 113, Birkhäuser, Basel, 2000, pp. 97–113.

[6] T. Hytonen and C. Perez, *Sharp weighted bounds involving A_∞*, Anal. PDE, **6**, (2013), no. 4, 777–818.

[7] R. F. Gundy and R. L. Wheeden, *Weighted integral inequalities for the nontangential maximal function, Lusin area integral, and Walsh-Paley series*, Studia Math. 49 (1973/74), 107–124.

[8] P. Ivanisvili, N. N. Osipov, D. M. Stolyarov, V. V. Vasyunin, P. B. Zatitskiy, *Bellman function for extremal problems in BMO*, Trans. Amer. Math. Soc. **368** (2016), 3415–3468

[9] M. T. Lacey, *An elementary proof of the A_2 bound*, Israel J. Math. **217** (2017), no. 1, 181–195.

[10] M. T. Lacey and K. Li, *On A_p–A_∞ type estimates for square functions*, Math. Z. **284** (2016), no. 3-4, 1211–1222.

[11] F. Nazarov, S. Petermichl, S. Treil, and A. Volberg, *Convex body domination and weighted estimates with matrix weights*, arXiv:1701.01907 [math.CA] (2017), 22pp.

[12] S. Petermichl and S. Pott, *An estimate for weighted Hilbert transform via square functions*, Trans. Amer. Math. Soc. **354** (2002), no. 4, 1699–1703.

[13] C. Thiele, S. Treil, and A. Volberg, *Weighted martingale multipliers in the non-homogeneous setting and outer measure spaces*, Adv. Math. **285** (2015), 1155–1188.

[14] S. Treil and A. Volberg, *Entropy conditions in two weight inequalities for singular integral operators*, Adv. Math. **301** (2016), 499–548.

[15] G. Wang, *Sharp square-function inequalities for conditionally symmetric martingales*, Trans. Amer. Math. Soc. **328** (1991), no. 1, 393–419.

[16] R. F. Gundy and R. L. Wheeden, *Weighted Integral Inequalities for the Nontangential Maximal function, Lusin area integral and Walsh Paley series*, Studia Math. **49**, (1973/74), 107–124.

[17] J. M. Wilson, *Weighted norm inequalities for the continuous square function*, Trans. Amer. Math. Soc. **314** (1989), no. 2, 661–692.

[18] J. M. Wilson, *L^p weighted norm inequalities for the square function, $0 < p < 2$*, Illinois J. Math. **33** (1989), Iss. 3, 361–366.

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

K. Domelevo: Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse, France
E-mail address: komla.domelevo@math.univ-toulouse.fr

P. Ivanisvili: Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA
E-mail address: ivanishvili.paata@gmail.com

S. Petermichl: Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse, France
E-mail address: stefanie.petermichl@gmail.com

S. Treil: Department of Mathematics, Brown University, Providence, RI 02912, USA
E-mail address: treil@math.brown.edu

A. Volberg: Department of Mathematics, Michigan State University, East Lansing, MI, 48824, USA
E-mail address: volberg@math.msu.edu