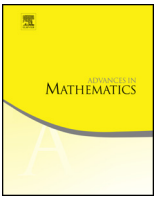
Advances in Mathematics 376 (2021) 107450

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim



"Small step" remodeling and counterexamples for weighted estimates with arbitrarily "smooth" weights ☆

S. Kakaroumpas *, S. Treil

Department of Mathematics, Brown University, Providence, RI 02912, USA

ARTICLE INFO

Article history:
Received 30 January 2019
Received in revised form 30 July 2020
Accepted 4 October 2020
Available online 24 October 2020
Communicated by Gilles Pisier

Keywords:
Calderón–Zygmund operators
Weighted estimates
Muckenhoupt characteristics
Poisson characteristics
Doubling constants
Dyadic filtration

ABSTRACT

For an A_p weight w the norm of the Hilbert Transform in $L^p(w)$, $1 < p < \infty$ is estimated by $[w]^s_{A_p}$, where $[w]_{A_p}$ is the A_p characteristic of the weight w and $s = \max(1, 1/(p-1))$; as simple examples with power weights show, these estimates are sharp.

A natural question to ask, is whether it is possible to improve the exponent s in the above estimate if one replaces the A_p characteristic by its "fattened" version, where the averages are replaced by Poisson-like averages. For power weights (for example with $p = 2$ and Poisson averages) one can see that there is indeed an improvement in the exponent: but is it true for general weights?

In this paper we show that the optimal exponent s remains the same by constructing counterexamples for arbitrarily "smooth" weights (in the sense that the doubling constant is arbitrarily close to 2), so the "fattened" A_p characteristic is equivalent to the classical one, and such that $\|T\|_{L^p(w)} \sim [w]^s_{A_p}$.

We use the ideas from the unpublished manuscript by F. Nazarov disproving Sarason's conjecture. We start from simple classical counterexamples for dyadic models, and then by using what we call "small step construction" we transform them into examples with weights that are arbitrarily dyadi-

☆ Supported in part by The National Science Foundation under the grants DMS-1600139, DMS-1856719.
* Corresponding author. *Current address:* Institut für Mathematik, Julius-Maximilians-Universität Würzburg, 97074 Würzburg, Germany.
E-mail addresses: spyridon.kakaroumpas@mathematik.uni-wuerzburg.de (S. Kakaroumpas), treil@math.brown.edu (S. Treil).

https://doi.org/10.1016/j.aim.2020.107450
0001-8708/© 2020 Elsevier Inc. All rights reserved.

cally smooth. F. Nazarov had used Bellman function method to prove the existence of such examples, but our construction gives a way to get such examples from the standard dyadic ones. We then use a modification of "remodeling", introduced by J. Bourgain and developed by F. Nazarov, to get from examples for dyadic models to examples for the Hilbert transform.

As an added bonus, we present a proof that the L^p analog of Sarason's conjecture is false for all p, $1 < p < \infty$.

© 2020 Elsevier Inc. All rights reserved.

Contents

Notation . . . 2
1. Introduction . . . 3
2. Preliminaries . . . 9
3. "Large step" examples . . . 15
4. "Small step" constructions . . . 18
5. Iterated remodeling . . . 28
6. The case of dyadic models . . . 35
7. The case of the Hilbert transform . . . 39
8. Appendix . . . 47
References . . . 51

Notation

$\mathbf{1}_E$	characteristic function of set E;
dx	integration with respect to Lebesgue measure;
$\|E\|$	d-dimensional Lebesgue measure of a measurable set $E \subseteq \mathbb{R}^d$;
$\langle f \rangle_E$	average with respect to Lebesgue measure, $\langle f \rangle_E := \frac{1}{\|E\|} \int_E f(x)dx$;
$L^p(w)$	weighted Lebesgue space, $\|f\|^p_{L^p(w)} := \int_{\mathbb{R}^d} \|f(x)\|^p w(x)dx$;
$\langle f, g \rangle$	linear duality, $\langle f, g \rangle = \int f(x)g(x)dx$;
$w(I)$	Lebesgue integral of a weight w over I, $w(I) := \int_I w(x)dx = \langle w \rangle_I \|I\|$;
p'	Hölder conjugate exponent to p, $1/p + 1/p' = 1$;
$\mathcal{D}$	family of all dyadic intervals in $\mathbb{R}$, or of all dyadic subintervals of $[0, 1)$;
$\mathcal{D}(I)$	family of all dyadic subintervals of a dyadic interval I, including I itself;
$\mathrm{ch}(I)$	family of all dyadic children of the dyadic interval I;
$\mathrm{ch}^k(I)$	family of all dyadic descendants of order k of the dyadic interval I, note that $\mathrm{ch}(I) = \mathrm{ch}^1(I)$;
$\mathrm{ch}^k(\mathscr{S})$	for a family $\mathscr{S}$ of dyadic intervals the collection $\mathrm{ch}^k(\mathscr{S})$ is defined as $\mathrm{ch}^k(\mathscr{S}) := \bigcup_{I \in \mathscr{S}} \mathrm{ch}^k(I)$, and $\mathrm{ch}(\mathscr{S}) = \mathrm{ch}^1(\mathscr{S})$;
I_-, I_+	left, respectively right half of interval I;
h_I	L^∞-normalized Haar function for interval I, $h_I := \mathbf{1}_{I_+} - \mathbf{1}_{I_-}$ (note the non-standard normalization!);

Δ_I martingale difference operator, $\Delta_I f := \sum_{I' \in \mathrm{ch}(I)} \langle f \rangle_{I'} \mathbf{1}_{I'} - \langle f \rangle_I \mathbf{1}_I$;

$\triangle_I f$ difference of averages, $\triangle_I f := \langle f \rangle_{I_+} - \langle f \rangle_I = \Big(\langle f \rangle_{I_+} - \langle f \rangle_{I_-} \Big)/2 = \langle f h_I \rangle_I$;

Notation $x \lesssim y$ means $x \le Cy$ with an absolute constant $C < \infty$, and $x \lesssim_{a,b,\dots} y$ means that C depends *only* on $a, b, \dots$; the notation $x \gtrsim y$ means $y \lesssim x$, and similarly for $x \gtrsim_{a,b,\dots} y$. We use $x \sim y$ if *both* $x \lesssim y$ and $x \gtrsim y$ hold, and $x \sim_{a,b\dots} y$ is defined similarly.

1. Introduction

This paper deals with sharp weighted estimates for classical operators in harmonic analysis. Our starting point is the famous Hunt–Muckenhoupt–Wheeden theorem [6], which says that the so-called Muckenhoupt A_p condition

$$\sup_I \left(\frac{1}{|I|} \int_I w(x)dx \right) \left(\frac{1}{|I|} \int_I w(x)^{-1/(p-1)} dx \right)^{p-1} =: [w]_{A_p} < \infty, \tag{1.1}$$

(the supremum is taken over all intervals $I \subset \mathbb{R}$) is necessary and sufficient for the Hilbert transform H,

$$Hf(x) := \text{p.v.} \int_{\mathbb{R}} \frac{f(y)}{x-y} dy,$$

to be a bounded operator on the weighted space $L^p(w)$ $(1 < p < \infty)$.

It is also well-known that condition (1.1) (with intervals replaced by cubes) is sufficient for the boundedness on weighted spaces of all Calderón–Zygmund operators in any number of dimensions, and it is also necessary for the boundedness on weighted spaces of "large" Calderón–Zygmund operators, like the Riesz transforms (see for instance [14]).

Remark. Recall that by a weight people usually understand a locally integrable non-negative function w, but to define A_p characteristic $[w]_{A_p}$ in (1.1) one needs to assume that w is positive a.e.

However, if we interpret $1/0$ as $+\infty$, then for a non-trivial weight w vanishing on a set of positive measure we have $[w]_{A_p} = \infty$, so the condition (1.1) fails for w. And it is easy to see that if a weight w vanishes on a set of positive measure, the Hilbert transform is not well defined on $L^p(w)$ (take a function f supported on the set where w vanishes), so we can treat the Hilbert transform as unbounded in this case.

So, one can say that the Hunt–Muckenhoupt–Wheeden theorem holds for arbitrary non-negative weights, if everything is interpreted the right way. However, to avoid confusing the reader with irrelevant technical details, we assume in this paper that a weight is always locally integrable a.e. positive function.

1.1. Sharp estimates

Qualitative results like the one mentioned above are usually easier to prove than quantitative counterparts. In fact, it had been an open problem for some time to find a sharp estimate of the norm of H (and other Calderón–Zygmund operators) over $L^p(w)$ in terms of the powers of the A_p characteristic $[w]_{A_p}$ defined in (1.1) above. It was proved by S. Petermichl in [12] that $\|H\|_{L^2(w)} \lesssim [w]_{A_2}$. She then proved the same estimate for the Riesz Transform, and after some results by different authors gradually expanding the class of operators for which such an estimate holds, the linear estimate $\|T\|_{L^2(w)} \lesssim_{T,d} [w]_{A_2}$ was established by T. Hytönen [7] (where d is the dimension of the underlying Euclidean space).

Using the method of Rubio De Francia extrapolation (see e.g. [3]), one then can show that for $p > 2$ the estimate $\|T\|_{L^p(w)} \lesssim_{T,p,d} [w]_{A_p}$ holds; by duality one finally gets the estimate $\|T\|_{L^p(w)} \lesssim_{T,p,d} [w]_{A_p}^{1/(p-1)}$ for $1 < p < 2$. Thus for $1 < p < \infty$ one can write

$$\|T\|_{L^p(w)} \lesssim_{T,p,d} [w]_{A_p}^{s}, \tag{1.2}$$

where

$$s = \max\{1, 1/(p-1))\}. \tag{1.3}$$

Note, that for the Hilbert transform the above estimate (1.2) is sharp. Namely, even before the upper bound for the Hilbert transform was proved by S. Petermichl [12], it had already been shown by S. Buckley [2] that given $p \in (1, \infty)$ one can find A_p weights w with arbitrarily large $[w]_{A_p}$ for which $\|H\|_{L^p(w)} \gtrsim_p [w]_{A_p}^{s}$ (with s given by (1.3)). It is also not hard to show that the estimate (1.2) is sharp for the Riesz transforms.

1.2. Considering "larger" characteristics

A reasonable attempt to lower the optimal exponent s given by (1.3) might involve considering "larger" variants of A_p characteristics where weights are not averaged over intervals (or cubes) as in (1.1), but rather integrated against kernels with slower decay. Such characteristics arise in fact naturally in many problems not directly related to sharp weighted estimates.

For instance, it was proved by the second author and A. Volberg in [15] for $p = 2$, and by F. Nazarov and the second author in [11] for general p, that the following "fattened" A_p condition[1] A_p^{fat}, $[w]_{A_p}^{\text{fat}} < \infty$, where

[1] In fact, in both [15] and [11] the estimates with matrix-valued weights were considered, and the Hunt–Muckenhoupt–Wheeden theorem for the matrix-valued weights was obtained. A matrix-valued analogue of the condition $[w]_{A_p}^{\text{fat}} < \infty$ was introduced there, and its necessity was proved. The necessity of the scalar condition follows immediately from the matrix-result, although just following the proofs from [15], [11] and not bothering with the non-commutativity of the matrix-valued case gives a very simple proof for the scalar situation.

$$[w]_{A_p}^{\mathrm{fat}} := \sup_{\lambda\in\mathbb{C}_+}\left(\int_{\mathbb{R}}\frac{(\mathrm{Im}(\lambda))^{p-1}}{|x-\lambda|^p}w(x)dx\right)\left(\int_{\mathbb{R}}\frac{(\mathrm{Im}(\lambda))^{p'-1}}{|x-\lambda|^{p'}}w(x)^{-1/(p-1)}dx\right)^{p-1}, \quad (1.4)$$

is necessary for the boundedness of the Hilbert transform on the weighted space $L^p(w)$; here, p' denotes the Hölder conjugate of p, $1/p+1/p'=1$. Note that for $p=2$, the integrals in (1.2) are just Poisson extensions of the weights w and w^{-1} (up to multiplicative constants), so one can think of $[w]_{A_p}^{\mathrm{fat}}$ as a "Poisson-like" A_p condition for any $1<p<\infty$. Motivation for considering such "Poisson-like" A_p conditions stems from the theory of Toeplitz operators, see for example [5, s. 7.9].

It is easy to see that $[w]_{A_p}\lesssim_p [w]_{A_p}^{\mathrm{fat}}$. Since the A_p condition is already sufficient for the boundedness of H on $L^p(w)$, it follows that the A_p condition and the "fattened" A_p condition A_p^{fat} are equivalent. However, simple examples involving power weights show that for every fixed p, the two characteristics themselves are not equivalent: for any fixed $1<p<\infty$, one can find A_p weights w with arbitrarily large quotient $[w]_{A_p}^{\mathrm{fat}}/[w]_{A_p}$. Moreover, it was shown in [15] that for $p=2$ the lower bound for the Hilbert Transform

$$\|H\|_{L^2(w)}\gtrsim \left([w]_{A_2}^{\mathrm{fat}}\right)^{1/2}$$

holds for *all* weights w.

So one could hope that a better estimate of the norm $\|T\|_{L^p(w)}$, and in particular of the norm $\|H\|_{L^p(w)}$, in terms of the "fattened" A_p characteristic $[w]_{A_p}^{\mathrm{fat}}$ in (1.4) is possible. One could even hope, for example, that the estimate $\|H\|_{L^2(w)}\lesssim \left([w]_{A_2}^{\mathrm{fat}}\right)^{1/2}$ holds. The main result of this paper destroys all such hopes: we show that for the Hilbert transform H there exist A_p weights w with arbitrarily large A_p characteristic $[w]_{A_p}^{\mathrm{fat}}$, such that $\|H\|_{L^p(w)}\gtrsim_p \left([w]_{A_p}^{\mathrm{fat}}\right)^s$, where s is given by (1.3).

1.2.1. "Heat" A_p characteristics

In many problems it is natural to consider other kernels besides "Poisson-like" ones. For example, S. Petermichl and A. Volberg [13] considered a "heat" A_p characteristic $(1<p<\infty)$ given by

$$[w]_{A_p}^{\mathrm{heat}} := \sup_{\substack{y\in\mathbb{R}^d\\ t\in(0,\infty)}}\left(\int_{\mathbb{R}^d}\frac{1}{t^{d/2}}e^{-|x-y|^2/t}w(x)dx\right)\left(\int_{\mathbb{R}^d}\frac{1}{t^{d/2}}e^{-|x-y|^2/t}w(x)^{-1/(p-1)}dx\right)^{p-1}. \quad (1.5)$$

It was shown in [13] that in sharp contrast to the "Poisson-like" case, the "heat" A_p characteristic in (1.5) is essentially the same as the usual Muckenhoupt A_p characteristic in (1.1), more precisely

$$[w]_{A_p}\sim_{d,p}[w]_{A_p}^{\mathrm{heat}} \qquad (1<p<\infty). \quad (1.6)$$

This fact for $d = 2$ was used [13] to establish sharp weighted estimates for the Ahlfors–Beurling operator, which allowed the authors to deduce that weakly quasiregular maps on the plane are quasiregular.

In view of (1.6) the problems considered in this paper are trivial for the "heat" A_p characteristic.

1.3. Weights and doubling constants

For a weight w on $\mathbb{R}$ we define its doubling constant D_w as

$$D_w := \sup_I w(2I)/w(I),$$

where the supremum is taken over all intervals I in $\mathbb{R}$. Here $2I$ is the interval with the same center as I of length $2|I|$, and slightly abusing the notation we write $w(I)$ for $\int_I w dx$.

It is easy to show that if the doubling constant of the weight w is bounded by $2 + \delta$ for sufficiently small δ, then we have uniformly over all $\lambda \in \mathbb{C}_+$ the estimate

$$\int_{\mathbb{R}} \frac{(\mathrm{Im}(\lambda))^{p-1}}{|x-\lambda|^p} w(x)dx \lesssim_p |I_\lambda|^{-1} \int_{I_\lambda} w(x)dx, \tag{1.7}$$

where I_λ is the interval $[\mathrm{Re}(\lambda) - \mathrm{Im}(\lambda), \mathrm{Re}(\lambda) + \mathrm{Im}(\lambda)]$. We emphasize that the particular function $(\mathrm{Im}\,\lambda)^{p-1}/|x-\lambda|^p$ in the left-hand side of (1.7) is of no importance here; any "reasonable" approximate identity on the real line can be used in its place.

Thus, if the doubling constants of the weights w and $\sigma = w^{-1/(p-1)}$ are bounded by $2 + \delta$ for sufficiently small δ, then the A_p characteristics $[w]_{A_p}$ and $[w]^{\mathrm{fat}}_{A_p}$ are equivalent in the sense of two sided estimate.

1.4. Main results

The main result of this paper is the following theorem.

Theorem 1.1. *Given $p \in (1, \infty)$, $M > 2$ and arbitrarily small $\delta > 0$, there exists an A_p weight w on $\mathbb{R}$ with $M \le [w]_{A_p} \le C(p)M$, such that the doubling constants of the weights w and $\sigma = w^{-1/(p-1)}$ are bounded by $2 + \delta$ and*

$$\|H\|_{L^p(w)} \ge c(p)M^s, \qquad s = \max\{1, 1/p - 1\}.$$

By the above discussion about the equivalence of A_p characteristics $[w]_{A_p}$ and $[w]^{\mathrm{fat}}_{A_p}$, we can see that Theorem 1.1 implies the following corollary.

Corollary 1.2. *Given $p \in (1,\infty)$, $M > 2$, there exists a weight w on $\mathbb{R}$ with $M \le [w]_{A_p}^{fat} \le C(p)M$, such that*

$$\|H\|_{L^p(w)} \ge c(p)M^s, \qquad s = \max\{1, 1/(p-1)\}.$$

1.4.1. Two weight estimates and Sarason's conjecture

One of the main technical tools used in this paper is inspired by the unpublished manuscript [9] by F. Nazarov, where he provided a counterexample to the so-called Sarason's conjecture. Let us briefly recall this conjecture.

It is natural to consider *two-weight* estimates for the Hilbert transform and other Calderón–Zygmund operators, i.e. to ask when they are bounded operators from $L^p(v)$ to $L^p(w)$ for potentially *different* weights v, w. It is easy to show that the two weight A_p condition

$$\sup_I \left(\frac{1}{|I|}\int_I w(x)dx\right)\left(\frac{1}{|I|}\int_I v(x)^{-1/(p-1)}dx\right)^{p-1} =: [w, v^{-1/(p-1)}]_{A_p} < \infty, \quad (1.8)$$

is necessary for the Hilbert transform to be a bounded operator from $L^p(v)$ to $L^p(w)$ $(1 < p < \infty)$. However, as simple examples show, this condition is not sufficient (for the reader's convenience we supply an example in Subsection 8.4 in the Appendix).

It had been shown long ago by the second author that the following "fattened" two weight A_p condition

$$\sup_{\lambda\in\mathbb{C}_+}\left(\int_{\mathbb{R}} \frac{(\mathrm{Im}(\lambda))^{p-1}}{|x-\lambda|^p} w(x)dx\right)\left(\int_{\mathbb{R}} \frac{(\mathrm{Im}(\lambda))^{p'-1}}{|x-\lambda|^{p'}} v(x)^{-1/(p-1)}dx\right)^{p-1} < \infty \quad (1.9)$$

is also necessary for the Hilbert transform to act boundedly from $L^p(v)$ to $L^p(w)$. Note, that unlike the one-weight case, the two-weight conditions (1.8) and (1.9) are not equivalent; simple examples can be easily constructed.

The Poisson averages are less localized than the averages over intervals, so D. Sarason hoped that for $p = 2$ the two weight Poisson A_2 condition (1.9) would capture correctly the "far" action of the Hilbert transform. In [5, s. 7.9] he conjectured that (for $p = 2$) the Poisson A_2 condition (1.9) is necessary and sufficient for the Hilbert transform to be a bounded operator from $L^2(v)$ to $L^2(w)$.[2]

This conjecture was disproved by F. Nazarov in [9]. In this paper we extend Nazarov's result to all $p \in (1,\infty)$ (not just $p = 2$). While our proof relies heavily on the machinery

[2] It is interesting that when D. Sarason was stating his conjecture he was not aware of the necessity of the two weight Poisson A_2 condition. The proof of necessity was presented to him by the second author, and this is exactly the proof presented (with attribution) in [5, s. 7.9].

The problem in [5, s. 7.9] was stated a bit different, but it was equivalent to the two weight estimate for the Hilbert transform. The proof of necessity was presented there only for $p = 2$, but the same proof works for all p, $1 < p < \infty$.

developed in [9], we introduce some crucial new ideas, allowing us to treat the case of $p \neq 2$. We should also mention that our counterexample is a "constructive" one; unlike [9] we are not using the Bellman function method.

We prove the following theorem:

Theorem 1.3. *Given $p \in (1, \infty)$, there exist weights w, v on $\mathbb{R}$ satisfying (1.9), such that the Hilbert transform is not a bounded operator acting from $L^p(v)$ to $L^p(w)$. In particular, this means that there exists $f \in L^p(v)$ such that $\|Hf\|_{L^p(w)} = \infty$.*

In light of the discussion in Section 1.3 the above theorem follows from the corresponding counterexample with "smooth" weights (i.e. weights with small doubling constants). Namely, we prove the following theorem, which implies the above Theorem 1.3.

Theorem 1.4. *Given $p \in (1, \infty)$ and arbitrarily small $\delta > 0$, there exist weights w, v on $\mathbb{R}$ satisfying (1.8), such that the doubling constants of the weights w and $\sigma = v^{-1/(p-1)}$ are bounded by $2+\delta$ and the Hilbert transform is not a bounded operator acting from $L^p(v)$ to $L^p(w)$. In particular, this means that there exists $f \in L^p(v)$ such that $\|Hf\|_{L^p(w)} = \infty$.*

1.4.2. A counterintuitive result

It is an easy exercise to construct a weight with a prescribed A_p characteristic. Moreover, one can find a weight taking only 2 values. What is more interesting, and is not completely clear, is that in fact one can find such a weight with doubling constant arbitrarily close to 2.

Proposition 1.5. *Let $p \in (1, \infty)$. Then, given $Q > 1$ and arbitrarily small $\varepsilon > 0$, there exists a weight w on $\mathbb{R}$ taking only 2 values, with $Q \leq [w]_{A_p} \leq c(p)Q$, such that the doubling constants of the weights w and $\sigma = w^{-1/(p-1)}$ are bounded by $2 + \varepsilon$.*

1.5. Plan of the paper

Our general strategy is as follows. We start with simple examples that give the desired lower bounds for dyadic (martingale) analogues of the Hilbert transform, in particular, for the so-called Haar shifts. These examples are simple ones, obtained as easy modifications of known examples; we call them the "large step" examples, to emphasize that we do not have any non-trivial bounds on the doubling constants of the weights involved. This is done in Section 3.

From these examples we construct in Section 4 the so-called "small step" examples, where we preserve the desired lower bounds, but can make the so-called *dyadic smoothness constant* (see the relevant definition in Subsection 2.3 below) of the weights as close to 1 as we want. We present a general construction that allows us to do so. This step is absent in [9], where the "small step" example is obtained implicitly via the Bellman function method.

The next step is to apply *remodeling*, introduced in [9], which serves two purposes. First, it allows us to get from weights with dyadic smoothness constants arbitrarily close to 1 to weights with doubling constants arbitrarily close to 2. And second (and equally important) it allows us to get from the lower bounds for Haar shifts to the lower bounds for the Hilbert transform, which we need. However, the original remodeling from [9] does not handle the one-weight situation well, since typically it gives a two-weight situation as its output. So to handle the one-weight situation we introduce the so-called *iterated remodeling*, that allows us to prove Theorem 1.1 (and so Corollary 1.2). The general method of iterated remodeling is presented in Section 5, while Subsection 7.1 contains the particular application for the Hilbert transform. Subsection 6.1 describes analogous examples in the (easier) cases of Haar multipliers and the dyadic Hardy–Littlewood maximal function. Moreover, Subsection 6.2 contains the counterintuitive result of Proposition 1.5, deduced as a byproduct of our general constructions.

Through a standard *direct sum of singularities* type construction, the family of examples for the Hilbert transform yields in Subsection 7.2 a counterexample to the L^p version of the Sarason's conjecture, (i.e. Theorem 1.4, and therefore Theorem 1.3), so we are done in the two-weight case as well.

The main constructions of this paper exploit the usual structure of a filtered probability space on the unit interval $[0, 1)$, and the fundamental correspondences between functions and martingales on the one hand, and martingales and random walks on graphs on the other hand. We briefly recall the relevant definitions and results in Subsections 2.4, 2.5 and 2.6.

Finally, in the Appendix (Section 8) we collect a few results used throughout the paper: probability theoretic results on random walks (Subsection 8.1), two remarks about "stopping on the lower hyperbola" (Subsection 8.2) and "getting only a little above the upper hyperbola" (Subsection 8.3), and we repeat the proofs of F. Nazarov's lemmas about Muckenhoupt characteristics and doubling constants from [9] (Subsection 8.5).

Acknowledgments. We are grateful to Alexander Barron for reading a draft of the manuscript and for pointing out typos and other obscurities, and to the anonymous referee for the valuable feedback.

2. Preliminaries

2.1. Symmetric "two weight" setup

In weighted estimates it is customary to rewrite a problem in a symmetric two-weight setup. For example, in an one-weight situation involving a weight w (Theorem 1.1) let us introduce an auxiliary weight $\sigma := w^{-1/(p-1)}$ (the reader should have noticed that it already appears in the statement of Theorem 1.1). If we denote $\widetilde{f} := \sigma^{-1} f$, so $f = \widetilde{f}\sigma$, then

$$\|\widetilde{f}\|_{L^p(\sigma)} = \|f\|_{L^p(w)} \qquad \text{and} \qquad Tf = T(\widetilde{f}\sigma),$$

for any linear operator T. Thus any weighted estimate of an operator T over $L^p(w)$ is equivalent to the estimate of the operator $\widetilde{f} \mapsto T(\widetilde{f}\sigma)$ acting from $L^p(\sigma)$ to $L^p(w)$; note that if T is an integral operator, then in the operator $f \mapsto T(f\sigma)$ integration is performed against the measure that defines the norm in the domain $L^p(\sigma)$.

To prove Theorem 1.1 one needs to find a non-zero $f \in L^p(w)$ such that $\|Hf\|_{L^p(w)} \geq c(p)\|f\|_{L^p(w)}$. This is equivalent to finding a non-zero $f \in L^p(\sigma)$ (we omit the tilde over f here) such that

$$\|H(f\sigma)\|_{L^p(w)} \geq c(p)M^s\|f\|_{L^p(\sigma)}; \tag{2.1}$$

here, recall, $M \leq [w]_{A_p} \leq C(p)M$, and $\sigma = w^{-1/(p-1)}$. The weights w and σ should have doubling constants as close to 2 as we want.

In a two-weight situation involving two weights w and v (Theorem 1.4) we denote $\sigma = v^{-1/(p-1)}$. To prove Theorem 1.4 we construct for arbitrarily large R weights σ and w with doubling constants arbitrarily close to 2 such that

$$\langle w\rangle_I \langle\sigma\rangle_I^{p-1} \leq C(p)$$

($C(p)$ does not depend on R) and a non-zero $f \in L^p(\sigma)$ such that

$$\|H(f\sigma)\|_{L^p(w)} \geq R\|f\|_{L^p(\sigma)}. \tag{2.2}$$

2.2. Dyadic intervals and martingale differences

For definiteness, by an interval we will always mean a half-open interval $[a, b)$. For an interval I we denote by I_+ and I_- its right and left halves respectively. The symbol h_I denotes the L^∞ normalized Haar function,

$$h_I = \mathbf{1}_{I_+} - \mathbf{1}_{I_-}. \tag{2.3}$$

We emphasize, that in this paper we always use the L^∞ normalized Haar functions.

We say that two intervals I, J in $\mathbb{R}$ are adjacent if $I \cap J = \varnothing$, and they have a common endpoint.

An interval I in $\mathbb{R}$ is called a dyadic interval if $I = [k2^n, (k+1)2^n)$ for some $n, k \in \mathbb{Z}$. We denote by $\mathcal{D}$ the family of all dyadic intervals in $\mathbb{R}$. For a dyadic interval I we denote by $\mathcal{D}(I)$ the collection of its dyadic subintervals (including I itself). When there is no danger of confusion, we will denote $\mathcal{D}([0, 1))$ by $\mathcal{D}$, abusing the notation. For all $I \in \mathcal{D}$, the number $-\log_2(|I|)$ will be called generation of the interval I. Moreover, for all $N \in \mathbb{N}$ and for all $I \in \mathcal{D}$, we denote by $\mathrm{ch}^N(I)$ (simply $\mathrm{ch}(I)$ if $N = 1$) the family of all dyadic subintervals of I of length $2^{-N}|I|$, and if $\mathcal{G}$ is a family of dyadic intervals, then we set $\mathrm{ch}^N(\mathcal{G}) := \bigcup_{I\in\mathcal{G}} \mathrm{ch}^N(I)$. Moreover, if $\mathcal{G}$ is a family of pairwise disjoint dyadic intervals then we denote

$$\mathbb{E}_{\mathcal{G}}[f] := \sum_{I\in\mathcal{G}} \langle f\rangle_I \mathbf{1}_I.$$

For all $f \in L^{\mathbf{1}}_{\mathrm{loc}}(\mathbb{R})$ and for all $I \in \mathcal{D}$, we denote by $\Delta_I f$ the martingale difference

$$\Delta_I f := \sum_{I'\in \mathrm{ch}(I)} \langle f\rangle_{I'} \mathbf{1}_{I'} - \langle f\rangle_I \mathbf{1}_I = \langle f\rangle_{I_+} 1_{I_+} + \langle f\rangle_{I_-} 1_{I_-} - \langle f\rangle_I \mathbf{1}_I,$$

and by ${}_{\vartriangle}I f$ the difference of averages (or Haar coefficient)

$${}_{\vartriangle}I f := \langle f\rangle_{I_+} - \langle f\rangle_I = \frac{\langle f\rangle_{I_+} - \langle f\rangle_{I_-}}{2} = \langle f h_I\rangle_I.$$

Notice that martingale differences and Haar coefficients are related by

$$\Delta_I f = ({}_{\vartriangle}I f) h_I.$$

2.3. Weights and doubling constants

Given weights w, σ on $\mathbb{R}$ and $p \in (1,\infty)$, we define the joint dyadic Muckenhoupt A_p characteristic of w, σ by

$$[w,\sigma]_{A_p,\mathcal{D}} := \sup_{I\in\mathcal{D}} \langle w\rangle_I \langle\sigma\rangle_I^{p-1}$$

and the dyadic Muckenhoupt characteristic of w by $[w]_{A_p,\mathcal{D}} := [w, w^{-1/(p-1)}]_{A_p,\mathcal{D}}$. Following [9, §1], we define the smoothness constant

$$S_w = \sup_I \max\left(\frac{\langle w\rangle_{I_-}}{\langle w\rangle_{I_+}}, \frac{\langle w\rangle_{I_+}}{\langle w\rangle_{I_-}}\right),$$

where the supremum is taken over all intervals I in $\mathbb{R}$, and the dyadic smoothness constant

$$S^{\mathrm{d}}_w = \sup_{I\in\mathcal{D}} \max\left(\frac{\langle w\rangle_{I_-}}{\langle w\rangle_{I_+}}, \frac{\langle w\rangle_{I_+}}{\langle w\rangle_{I_-}}\right).$$

It is easy to see that $D_w \le S_w + 1$. Note also that $1 \le S^{\mathrm{d}}_w \le S_w$. Moreover, as in [9, §6], we define the strong dyadic smoothness constant

$$S^{\mathrm{sd}}_w = \sup_{I,J} \frac{\langle w\rangle_I}{\langle w\rangle_J},$$

where the supremum is taken over all adjacent intervals $I, J \in \mathcal{D}$ with $|I| = |J|$. Obviously $S^{\mathrm{sd}}_w \ge S^{\mathrm{d}}_w$. Of course all these definitions can be given over $[0,1)$, and we will use the

same notation as above for Muckenhoupt characteristics and smoothness constants over $[0,1)$ (note that local integrability over $[0,1)$ means here integrability over $[0,1)$).

It turns out that the strong dyadic smoothness constant can provide some control over the smoothness constant, and the dyadic Muckenhoupt characteristic over the full Muckenhoupt characteristic, provided the strong dyadic smoothness constant is sufficiently close to 1.

Lemma 2.1. *(F. Nazarov, [9, §6]) For all $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$, such that for all weights w on $\mathbb{R}$ with $S_w^{sd} \leq 1+\delta$ there holds $S_w \leq 1+\varepsilon$.*

Lemma 2.2. *(F. Nazarov, [9, §11]) For all $p \in (1,\infty)$, there exists $\delta = \delta(p) > 0$, such that for all weights w, σ on $\mathbb{R}$ with $[w,\sigma]_{A_p,\mathcal{D}} < \infty$ and $S_w^{sd}, S_\sigma^{sd} \leq 1+\delta$ there holds $[w,\sigma]_{A_p} \leq (5/4)[w,\sigma]_{A_p,\mathcal{D}}$.*

For reasons of completeness, we give the proofs of both these lemmas in Subsection 8.5 in the Appendix. In this paper, the phrase "smoothness of weights" will always refer to the above smoothness constants.

So we see that in order to dominate Muckenhoupt characteristics and doubling constants, it suffices to dominate strong dyadic smoothness constants and dyadic Muckenhoupt characteristics. We will see in Section 5 that F. Nazarov's method of remodeling will allow us to dominate strong dyadic smoothness constants by dyadic smoothness constants.

2.4. Dyadic filtration

For $n = 0, 1, 2, \ldots$, set

$$\mathcal{D}_n = \{I \in \mathcal{D}([0,1)) : \ |I| = 2^{-n}\},$$

and let $\mathcal{F}_n$ be the σ-algebra of subsets of $[0,1)$ generated by the family $\mathcal{D}_n$, i.e. the smallest σ-algebra of subsets of $[0,1)$ containing $\mathcal{D}_n$. Clearly $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, for all $n = 0, 1, 2, \ldots$, so the sequence $\mathbb{F} := (\mathcal{F}_n)_{n=0}^{\infty}$ of σ-algebras is a filtration on $[0,1)$ (sometimes called the dyadic filtration). Notice that the Borel σ-algebra $\mathcal{F}$ of $[0,1)$ is the smallest σ-algebra containing all $\mathcal{F}_n$, or equivalently, the σ-algebra generated by the family $\bigcup_{n=0}^{\infty} \mathcal{D}_n$.

Taking for the probability measure $\mathbb{P}$ the Lebesgue measure on $[0,1)$, we can see that $([0,1), \mathcal{F}, \mathbb{P}, \mathbb{F})$ is a filtered probability space. Denote by $\mathbb{E}_n$ the conditional expectation with respect to the σ-algebra $\mathcal{F}_n$, $\mathbb{E}_n[f] = \mathbb{E}(f|\mathcal{F}_n)$. The operator $\mathbb{E}_n$ admits a simple formula

$$\mathbb{E}_n[f] = \sum_{I \in \mathcal{D}_n} \langle f \rangle_I \mathbf{1}_I.$$

We will use the symbol $\mathbb{E}$ for the expectation operator, $\mathbb{E}f := \mathbb{E}_0 f = \langle f \rangle_{[0,1)} \mathbf{1}_{[0,1)}$.

Recall that a sequence $(X_n)_{n=0}^{\infty}$ of integrable functions on a filtered probability space is called a martingale if X_n is $\mathcal{F}_n$-measurable and

$$\mathbb{E}_n[X_{n+1}] = X_n,$$

for all $n = 0, 1, 2, \ldots$. In the sequel, all martingales on $[0, 1)$ will always be considered with respect to the dyadic filtration (and called then just dyadic martingales).

Note that every dyadic interval can be given the structure of a filtered probability space by simply translating and rescaling the unit interval.

2.5. Functions and martingales

A function $f \in L^1([0, 1); \mathbb{R}^N)$ naturally induces an $\mathbb{R}^N$-valued martingale $X = (X_n)_{n=0}^{\infty}$ on $[0, 1)$,

$$X_n = \mathbb{E}_n f = \sum_{I \in \mathcal{D}_n} \langle f \rangle_I \mathbf{1}_I, \qquad n = 0, 1, 2, \ldots.$$

Note, that not all martingales are induced by a function, only the so-called *uniformly integrable* ones. However, in this paper we will be considering only uniformly bounded martingales, which are trivially uniformly integrable, and so are always induced by a function.

If a martingale X is induced by a function f, then f can be easily restored from X, namely $X_n \to f$ a.e. and in L^1; for uniformly bounded martingales we have, in fact, convergence in all L^p, $1 \le p < \infty$.

It turns out that in many problems of harmonic analysis it is more convenient to work not with a function, but with the induced martingale. In our context that means that we keep track of averages of functions, instead of the functions themselves. In our examples, we deal with functions w, σ, f, g, and we are keeping track of the averages of functions $w, \sigma, \mathbf{f} =: f\sigma, \mathbf{g} =: gw$ (then $f = \mathbf{f}/\sigma$ and $g = \mathbf{g}/w$).

2.6. Martingales and random walks

Let $X = (X_n)_{n=0}^{\infty}$ be an $\mathbb{R}^N$-valued martingale on $[0, 1)$. For $I \in \mathcal{D}_n$ the function X_n is constant on I; we denote by $\langle X \rangle_I$ its constant value there. Note that if the martingale X is induced by a function, which we, slightly abusing the notation, also denote by X, then $\langle X \rangle_I$ as defined above is indeed the average of the function X. It is easy to see that

$$\langle X \rangle_I = \frac{\langle X \rangle_{I_-} + \langle X \rangle_{I_+}}{2}, \qquad \forall I \in \mathcal{D}. \tag{2.4}$$

In the language of [4, Subsection 5.1] the above identity says that the family $\{\langle X \rangle_I\}_{I \in \mathcal{D}}$ has "martingale dynamics".

We also define the difference of averages (or Haar coefficient) $\triangle_I X$,

$$\triangle_I X := \langle X\rangle_{I_+} - \langle X\rangle_I = \frac{\langle X\rangle_{I_+} - \langle X\rangle_{I_-}}{2}, \qquad I \in \mathcal{D}.$$

Again, if, slightly abusing the notation, we denote by X the function inducing the uniformly bounded martingale X, then the two definitions of $\triangle_I X$ are consistent.

The dyadic martingale X can be interpreted as a random walk on an image of a binary tree; in what follows we will call this image the *graph* of X.

To describe this random walk, notice that the collection $\mathcal{D}$ of dyadic intervals carries a natural structure of the full binary tree, with vertices being the dyadic intervals, and the edges connecting an interval with its two children.

The collection $\mathcal{D}$ of dyadic intervals can be naturally interpreted as the standard random walk on the full dyadic tree, where one moves from a vertex I to each of its children with probability $1/2$. Each point $x \in [0,1)$ represents a trajectory on the full binary tree $\mathcal{D}$, that at the time n it is at the unique $I \in \mathcal{D}_n$ containing x.

The martingale X naturally induces a map from the dyadic tree $\mathcal{D}$ to $\mathbb{R}^N$, where the vertex corresponding to $I \in \mathcal{D}$ goes to the point $\langle X\rangle_I \in \mathbb{R}^N$, and the edges go to straight line segments connecting the corresponding points; we will call this image the *graph* of X. The random walk on the dyadic tree $\mathcal{D}$ is then mapped to the random walk on the graph of X, that moves from a point $\langle X\rangle_I$ by the steps $\pm\triangle_I X$ with equal probabilities $1/2$.

In view of the martingale dynamics identity (2.4) above, $\langle X\rangle_I$ always occupies the midpoint of the straight line segment connecting $\langle X\rangle_{I_-}$ and $\langle X\rangle_{I_+}$; we will say in what follows that this segment corresponds to the interval $I \in \mathcal{D}$.

The interpretation of dyadic martingales as random walks on images of a binary tree gives helpful intuition into the constructions we are using. While it is not required for the formal construction, we feel that it could help the reader to understand and visualize what is going on.

In our examples, we deal with functions w, σ, f, g, where $w = \sigma^{-1/(p-1)}$ for some $1 < p < \infty$, and random walks correspond to the martingales induced by the functions $w, \sigma, \mathbf{f} := f\sigma, \mathbf{g} := gw$. Our transforms will be applied to the functions $w, \sigma, \mathbf{f}, \mathbf{g}$, to produce functions $\widetilde{w}, \widetilde{\sigma}, \widetilde{\mathbf{f}}, \widetilde{\mathbf{g}}$ respectively. The random walk corresponding to the martingale induced by the function (w, σ) terminates with probability 1 on the hyperbola given in the uv-plane by $uv^{p-1} = 1$, because $w\sigma^{p-1} = 1$ a.e. on $[0,1)$. Our transforms will need to guarantee that the new weights $\widetilde{w}, \widetilde{\sigma}$ we get continue to satisfy this relation. As we will see, on the level of weights our transforms will amount to composition with measure-preserving transformations, and therefore such relations will be automatically preserved. In addition, we will see that the relevant weighted norms $\|\widetilde{f}\|_{L^p(\widetilde{\sigma})}$, $\|\widetilde{g}\|_{L^{p'}(\widetilde{w})}$ are not larger (up to constants depending only on p) than $\|f\|_{L^p(\sigma)}$, $\|g\|_{L^{p'}(w)}$ respectively, where $\widetilde{f} = \widetilde{\mathbf{f}}/\widetilde{\sigma}$ and $\widetilde{g} = \widetilde{\mathbf{g}}/\widetilde{w}$.

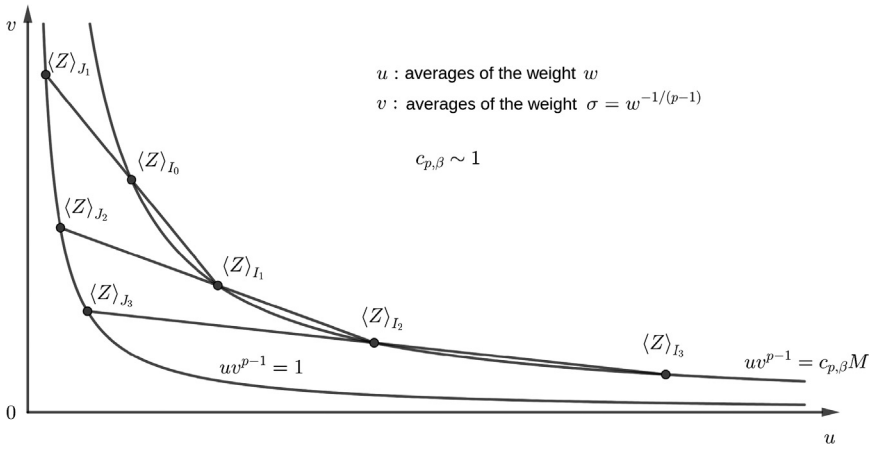


Fig. 1. Random walk in the uv-plane corresponding to the pair of weights (w, σ).

3. "Large step" examples

We construct in this section "large step" examples for the Haar multiplier, and for a special type of Haar shift, defined in Subsection 3.2.

Let $p \in (1, \infty)$ and $M > 2$. Set $\beta = 1 - \frac{1}{2Me} \in \left(\frac{1}{2}, 1\right)$. Set $I_0 = [0, 1)$ and $I_n = \left[0, \frac{1}{2^n}\right)$, $J_n = \left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)$, for all $n = 1, 2, \ldots$. Consider the functions w, σ on $[0, 1)$ given by

$$w = \sum_{n=1}^{\infty} 2^{n\beta} 1_{J_n}, \qquad \sigma = \sum_{n=1}^{\infty} 2^{-n\beta/(p-1)} 1_{J_n}.$$

Then, w, σ are weights on $[0, 1)$ with $\sigma = w^{-1/(p-1)}$. Note that $w([0, 1)) \sim M$ and $\sigma([0, 1)) \sim_p 1$. Notice that $x^{-\beta} \leq w(x) \leq 2^{\beta} x^{-\beta}$ and $2^{-\beta/(p-1)} x^{\beta/(p-1)} \leq \sigma(x) \leq x^{\beta/(p-1)}$, for all $x \in (0, 1)$. Then, direct computation shows that

$$M \leq 2^{-\beta} \frac{(1-\beta)^{-1}}{e} \leq \langle w \rangle_{I_n} \langle \sigma \rangle_{I_n}{}^{p-1} \leq 2^{\beta} (1-\beta)^{-1} \leq 4Me, \qquad \forall n = 0, 1, 2, \ldots.$$

It follows that $M \leq [w]_{A_p, \mathcal{D}} \leq 4Me$. Direct computation gives also $\vartriangle_{I_n} w < 0$ and $-\vartriangle_{I_n} w \sim (1-\beta)^{-1} 2^{n\beta}$, for all $n = 0, 1, 2, \ldots$.

Consider the uniformly integrable real-valued martingales X, Y induced by w, σ respectively. Note that by a very easy application of Jensen's inequality as in [10, Lemma 4.1] we have $X_n Y_n^{p-1} \geq 1$, for all $n = 0, 1, 2 \ldots$. Also note that the graph of the martingale $Z = (X, Y)$ consists of the straight line segments connecting $\langle Z \rangle_{J_n}$ and $\langle Z \rangle_{I_n}$, for $n = 1, 2, \ldots$, see Fig. 1 (the constant $c_{p,\beta}$ in Fig. 1 satisfies $1 \leq c_{p,\beta} \leq 4e$).

Notice moreover that $S_w^{\mathrm{d}} \sim (1-\beta)^{-1} \sim M$, therefore we have no control over the dyadic smoothness constant of w.

We will now truncate the weights w, σ. We have

$$\sum_{n=0}^{\infty} 2^{n(\beta-1)} = \frac{1}{1-2^{\beta-1}} \gtrsim (1-\beta)^{-1} = 2Me.$$

Therefore, there exists a positive integer $N = N_M$ greater than 1, such that

$$\sum_{n=0}^{N} 2^{n(\beta-1)} \gtrsim M.$$

The following lemma, whose proof is given in Subsection 8.2 of the appendix, implies that there exist $a_1, a_2, b_1, b_2 > 0$ such that $(a_1+a_2)/2 = \langle w \rangle_{I_{N+1}}$, $(b_1+b_2)/2 = \langle \sigma \rangle_{I_{N+1}}$ and $a_1 b_1^{p-1} = a_2 b_2^{p-1} = 1$.

Lemma 3.1. *Let $x, y > 0$ be arbitrary, such that $xy^{p-1} \geq 1$. Then, there exist $a_1, b_1, a_2, b_2 > 0$ with $a_2 \leq x \leq a_1$ and $b_1 \leq y \leq b_2$, such that $a_1 b_1^{p-1} = a_2 b_2^{p-1} = 1$ and $x = \frac{a_1+a_2}{2}$, $y = \frac{b_1+b_2}{2}$.*

Without loss of generality, we may assume that $a_1 < a_2$. Consider the bounded weights

$$w' = \sum_{n=1}^{N+1} 2^{n\beta} 1_{J_n} + a_1 1_{J_{N+2}} + a_2 1_{I_{N+2}}, \qquad \sigma' = \sum_{n=1}^{N+1} 2^{-n\beta/(p-1)} 1_{J_n} + b_1 1_{J_{N+2}} + b_2 1_{I_{N+2}}$$

on $[0,1)$. Notice that ${}_{\triangle I_{N+1}} w' = (a_1 - a_2)/2 < 0$. In what follows, we abuse the notation denoting w', σ' by w, σ respectively.

3.1. Example for the Haar multiplier

For any choice of signs $\varepsilon = (\varepsilon_I)_{I \in \mathcal{D}}$ denote by T_ε the Haar multiplier on $[0,1)$ corresponding to ε, i.e. T_ε acts on functions $f \in L^2([0,1))$ via

$$T_\varepsilon(f) = \sum_{I \in \mathcal{D}} \varepsilon_I ({}_{\triangle I} f) h_I.$$

Consider the function $\mathbf{f}$ on $[0,1)$ given by

$$\mathbf{f} = \sum_{n=1}^{\infty} (-1)^{n-1} 1_{J_n}.$$

Direct computation gives that for all $I \in \mathcal{D}$, we have ${}_{\triangle I}\mathbf{f} \neq 0$ if and only if $I = I_n$ for some $n \in \mathbb{N}$, in which case ${}_{\triangle I}\mathbf{f} = \frac{2(-1)^{n+1}}{3}$. Consider also the function $\mathbf{g} = -w$ on $[0,1)$. Consider the functions $f = \mathbf{f}/\sigma$, $g = \mathbf{g}/w$ on $(0,1)$. Then

$$\|f\|^p_{L^p(\sigma)} = w([0,1)) \sim M, \quad \|g\|^{p'}_{L^{p'}(w)} = w([0,1)) \sim M.$$

Moreover, we have

$$\sup_{\varepsilon\in\mathcal{E}} |\langle T_\varepsilon(f\sigma), gw\rangle| = \sup_{\varepsilon\in\mathcal{E}} \left|\sum_{I\in\mathcal{D}} \varepsilon_I |I| (\triangle_I \mathbf{f})(\triangle_I \mathbf{g})\right| = \sum_{I\in\mathcal{D}} |I| \cdot |\triangle_I \mathbf{f}| \cdot |\triangle_I w|$$
$$\geq \sum_{n=0}^{N} |I_n| \cdot |\triangle_{I_n} \mathbf{f}| \cdot |\triangle_{I_n} w| \sim (1-\beta)^{-1} \sum_{n=0}^{N} 2^{n(\beta-1)} \gtrsim (1-\beta)^{-1} M \sim M^2,$$

where $\mathcal{E}$ is the set of all choices of signs $\varepsilon = (\varepsilon_I)_{I\in\mathcal{D}}$. It follows that

$$\sup_{\varepsilon\in\mathcal{E}} \frac{|\langle T_\varepsilon(f\sigma), gw\rangle|}{\|f\|_{L^p(\sigma)} \|g\|_{L^{p'}(w)}} \gtrsim_p \frac{M^2}{M^{1/p} M^{1/p'}} = M.$$

3.2. Example for a special type of Haar shift

Let T be the Haar shift on $[0,1)$ acting on functions $f \in L^2([0,1))$ by

$$Tf = 2 \sum_{I\in\mathcal{D}} (\triangle_I f)(h_{I_+} - h_{I_-}).$$

Then, we have

$$\langle Tf, g\rangle = \sum_{I\in\mathcal{D}} |I| (\triangle_I f)(\triangle_{I_+} g - \triangle_{I_-} g),$$

for all $f, g \in L^2([0,1))$.

Consider the function $\mathbf{f}$ on $[0,1)$ given by

$$\mathbf{f} = \sum_{n=1}^{\infty} h_{J_n}.$$

Notice that $|\mathbf{f}| \leq 1$. It is obvious that for all $I \in \mathcal{D}$, we have $\triangle_I \mathbf{f} \neq 0$ if and only if $I = J_n$ for some positive integer n, in which case $\triangle_I \mathbf{f} = 1 > 0$. Consider also the function $\mathbf{g} = -w$ on $[0,1)$. Consider the functions $f = \mathbf{f}/\sigma$, $g = \mathbf{g}/w$ on $[0,1)$. We have

$$\|f\|^p_{L^p(\sigma)} = \left\|\frac{1}{\sigma}\right\|^p_{L^p(\sigma)} = w([0,1)) \sim M, \quad \|g\|^{p'}_{L^{p'}(w)} = w([0,1)) \sim M.$$

Moreover, we have

$$\langle f\sigma, T(gw)\rangle = \langle \mathbf{f}, T(\mathbf{g})\rangle \geq \sum_{n=0}^{N} |I_n| (\triangle_{I_n} \mathbf{g})(\triangle_{J_{n+1}} \mathbf{f}) \sim \sum_{n=0}^{N} (1-\beta)^{-1} 2^{n(\beta-1)} \gtrsim M^2.$$

It follows that

$$\frac{\langle f\sigma, T(gw)\rangle}{\|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}} \gtrsim_p \frac{M^2}{M^{1/p}M^{1/p'}} = M.$$

4. "Small step" constructions

We describe in this section different variants of "small step" constructions, that allow us to get from the examples constructed above in Section 3 to examples with dyadic smoothness constants arbitrarily close to 1.

We fix the following notation: for all intervals J, K in $\mathbb{R}$, we denote by $\psi_{J,K}$ the unique orientation-preserving affine transformation mapping J onto K.

4.1. A warmup: the "small step" construction for the Haar multiplier

Let $p \in (1,\infty)$ and $M > 2$. Recall that in Subsection 3.1 we constructed bounded weights w, σ on $[0,1)$ with $\sigma = w^{-1/(p-1)}$, such that

$$M \le w([0,1))\sigma([0,1))^{p-1}, \quad [w]_{A_p,\mathcal{D}} \le 4Me, \quad w([0,1)) \sim M, \quad \sigma([0,1)) \sim_p 1,$$

and non-zero bounded functions $f \in L^p(w)$, $g \in L^{p'}(\sigma)$ such that

$$\sup_{\varepsilon\in\mathcal{E}} |\langle T_\varepsilon(f\sigma), gw\rangle| = \sum_{I\in\mathcal{D}} |I| \cdot |\vartriangle_I \mathbf{f}| \cdot |\vartriangle_I \mathbf{g}| \ge c(p)\|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}, \tag{4.1}$$

where $\mathbf{f} := f\sigma$ and $\mathbf{g} := gw$. Recall that in this example we do not have any control over the dyadic smoothness constants S_w^{d} and S_σ^{d} of the weights w and σ.

Based on this example we want to construct weights $\widetilde{w}, \widetilde{\sigma}$ with $\widetilde{\sigma} = \widetilde{w}^{-1/(p-1)}$, and non-zero functions $\widetilde{f} \in L^p(\widetilde{w})$, $\widetilde{g} \in L^{p'}(\widetilde{\sigma})$ such that (4.1) holds with $\widetilde{f}$, $\widetilde{g}$, $\widetilde{w}$, $\widetilde{\sigma}$ in place of f, g, w, σ (with another constant $c(p)$); and what is essential, that the dyadic smoothness constants of the new weights are as close to 1 as we want.

As we will see, in our construction we will keep track of the averages and martingale differences of the weight w, σ and of the functions $\mathbf{f}$ and $\mathbf{g}$, and their counterparts with tildes.

4.1.1. A general "small step" construction

We begin by describing a "small step" construction that does not exploit any intricacies of the particular "large step" example for Haar multipliers.

Let us first give an informal description. Let X be an $\mathbb{R}^N$-valued martingale on $[0,1)$. Consider the *graph* of the martingale X, see Subsection 2.6. Recall, that the segment of the graph, corresponding to an interval $J \in \mathcal{D}$ is a straight line segment, connecting points $\langle X\rangle_{J_-}$ and $\langle X\rangle_{J_+}$; note that $\langle X\rangle_J$ is the midpoint of this segment.

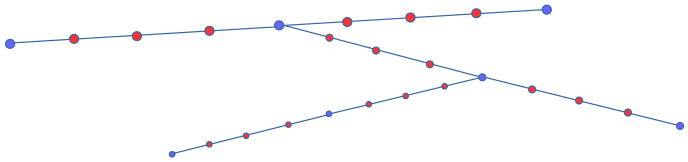


Fig. 2. Dividing the segments of the graph of X. (For interpretation of the colors in the figure, the reader is referred to the web version of this article.)

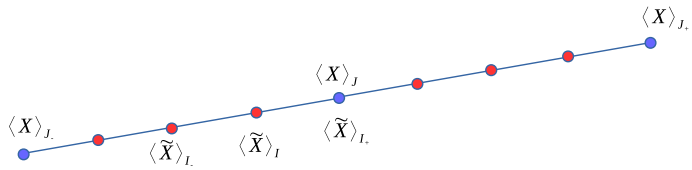


Fig. 3. Random walk on the new graph. (For interpretation of the colors in the figure, the reader is referred to the web version of this article.)

Take a sufficiently large positive integer d. We divide each of the segments of the graph of X in $2d$ parts, so that we get a new graph containing the vertices of the old graph, along with several new vertices, $2\cdot(d-1)$ in number, on each segment, see Fig. 2, where new points are marked in red.

Let us describe a new random walk on the new graph, which can be thought of as a "small step" version of the random walk corresponding to the original martingale, producing a new martingale $\widetilde{X}$.

As in the original random walk, we start from the average $\langle X\rangle_{[0,1)}$, which, recall, is the midpoint of the segment corresponding to $[0,1)$. From each point $\langle X\rangle_J$ we perform a "small step" random walk of order d along the segment corresponding to J, moving by $\pm\vartriangle_J X/d$ with probability $1/2$. Thus, from each point of the new graph, we move with equal probability $1/2$ to one of the two immediately closest points (of the new graph) on the corresponding segment (see Fig. 3). When we reach one of the two endpoints of this segment, we get into a new segment, and we repeat this procedure along the new segment.

Let us now make all this formal. In our case the martingale is always a uniformly bounded one, induced by a function $F\in L^\infty([0,1);\mathbb{R}^N)$; usually in our situation $N=4$ and $F=(w,\sigma,\mathbf{f},\mathbf{g})$. The construction will be described in terms of the function F, so no deep knowledge of probability is required, although the above probabilistic description could help the reader to understand what is going on.

Given a dyadic subinterval I of $[0,1)$, we define the family $\mathscr{S}(I)$ of stopping intervals for I as the family of all maximal dyadic subintervals J of I such that

$$\left|\sum_{\substack{I'\in\mathcal{D}(I)\\ I'\supsetneq J}} h_{I'}\right| = d, \tag{4.2}$$

and we also define the subset $\mathscr{S}_+(I)$ as the family of all intervals J in $\mathscr{S}(I)$ for which the sum in (4.2) is equal to d, and similarly we define $\mathscr{S}_-(I)$. Coupled with a translation and rescaling invariance lemma, part (i) of the following lemma implies that the family $\mathscr{S}(I)$ forms a partition (up to a Borel set of zero measure) of I, and part (ii) of it implies that $\bigcup \mathscr{S}_+(I), \bigcup \mathscr{S}_+(I)$ have both measure equal to $|I|/2$.

Lemma 4.1. *Consider the sequence $(r_n)_{n=1}^{\infty}$ of Rademacher functions on $[0,1)$, i.e.*

$$r_n := \sum_{I \in \mathcal{D}_{n-1}} h_I, \qquad n = 1, 2, \dots.$$

Set $S_0 = 0$ and $S_n = \sum_{k=1}^{n} r_k$, for all $n = 1, 2, \dots$. Let $a, b \geq 0$, not both of them equal to 0. Consider the stopping times τ^1, τ^2, τ given by

$$\tau^1 := \inf\{n \in \mathbb{N} : \ S_n = b\}, \qquad \tau^2 := \inf\{n \in \mathbb{N} : \ S_n = -a\}, \qquad \tau := \min(\tau^1, \tau^2).$$

(i) There holds $\tau^1 < \infty$ and $\tau^2 < \infty$ a.e. on $[0,1)$.
(ii) There holds $\mathbb{P}(\tau = \tau^1) = \frac{a}{a+b}$ and $\mathbb{P}(\tau = \tau^2) = \frac{b}{a+b}$.

The proof of the lemma is given in Subsection 8.1 of the Appendix.

The transformation we describe here acts on functions in $L^\infty(I)$ as follows. Let $G \in L^\infty(I; \mathbb{R}^N)$. Then, we define the function $R_I G := G \circ \psi_I$, where $\psi_I : I \to I$ is given by

$$\psi_I(x) = \begin{cases} \psi_{J, I_-}(x), & \text{if } x \text{ belongs to some } J \in \mathscr{S}_-(I) \\ \psi_{J, I_+}(x), & \text{if } x \text{ belongs to some } J \in \mathscr{S}_+(I) \end{cases}, \text{ for almost every } x \in I. \quad (4.3)$$

It is clear that $\psi_I : I \to I$ is a measure-preserving transformation.

The "small step" transform described here is obtained though iterating the above transform in every stopping interval. Namely, we first apply the above construction on the function F, along the interval $[0,1)$. We thus obtain a function $R_{[0,1)}F \in L^\infty([0,1); \mathbb{R}^N)$. Then, we apply the above transform on the function $(R_{[0,1)}F)|_I$ along the interval I, producing new stopping intervals, for all $I \in \mathscr{S}([0,1))$, and afterwards we repeat this along every stopping interval that will have come up, etc. Therefore, after this process has been completed we will have obtained a new function $\tilde{F} \in L^\infty([0,1); \mathbb{R}^N)$.

It is important to note that in fact this transform (called in what follows "small step" transform of order d) amounts just to a composition of limiting functions with a certain measure-preserving transformation (so in particular, it does not matter whether we apply it to a martingale as a whole or to each of its coordinates separately). Indeed, it is clear that $\tilde{F} = F \circ \Phi$, where $\Phi : [0,1) \to [0,1)$ is the measure-preserving transformation given at almost every point of $[0,1)$ as the composition of all the measure-preserving transformations $\psi_I : I \to I$, where I runs over $[0,1)$ and all stopping intervals containing that point (note that the order of composition respects inclusion of dyadic intervals).

We now specialize to the case $N = 4$ and $F = (w, \sigma, \mathbf{f}, \mathbf{g})$. We write then $\tilde{F} = (\widetilde{w}, \widetilde{\sigma}, \widetilde{\mathbf{f}}, \widetilde{\mathbf{g}})$, where tilde denotes just composition with the measure preserving transformation Φ. In particular, $\tilde{w}, \tilde{\sigma}$ are weights on $[0,1)$ with $\widetilde{w}\widetilde{\sigma}^{p-1} = 1$ a.e. on $[0,1)$.

4.1.2. Getting the damage

We first show that the "small step" transform preserves "damage" for Haar multipliers.

Lemma 4.2. *Let the functions $\mathbf{f}, \mathbf{g}, \widetilde{\mathbf{f}}, \widetilde{\mathbf{g}}$ be as above. There holds*

$$\sum_{I\in\mathcal{D}} |I| \cdot |\vartriangle_I \widetilde{\mathbf{f}}| \cdot |\vartriangle_I \widetilde{\mathbf{g}}| = \sum_{J\in\mathcal{D}} |J| \cdot |\vartriangle_J \mathbf{f}| \cdot |\vartriangle_J \mathbf{g}|.$$

Proof. First of all, it is immediate by translation and rescaling invariance that

$$\sum_{J\in\mathscr{S}(I_0)} \sum_{K\in\mathcal{D}(J)} |\vartriangle_K \hat{\mathbf{f}}| \cdot |\vartriangle_K \hat{\mathbf{g}}| \cdot |K| = \sum_{\substack{J\in\mathcal{D}(I_0)\\ J\neq I_0}} |\vartriangle_J \mathbf{f}| \cdot |\vartriangle_J \mathbf{g}| \cdot |J|,$$

where $I_0 = [0,1)$ and $\hat{\mathbf{f}} := R_{I_0}\mathbf{f},\ \hat{\mathbf{g}} := R_{I_0}\mathbf{g}$. Therefore, since the transform is given by iteration of the same fundamental transform over $[0,1)$ and all stopping intervals, up to translation and rescaling, it suffices only to verify that

$$\sum_{K\in\mathcal{D}(I_0)\setminus\left(\bigcup_{J\in\mathscr{S}(I_0)}\mathcal{D}(J)\right)} |\vartriangle_K \hat{\mathbf{f}}| \cdot |\vartriangle_K \hat{\mathbf{g}}| \cdot |K| = |\vartriangle_{I_0} \mathbf{f}| \cdot |\vartriangle_{I_0} \mathbf{g}| \cdot |I_0|. \tag{4.4}$$

It is easy to verify that

$$\vartriangle_K \hat{\mathbf{f}} = \frac{1}{d} \vartriangle_{I_0} \mathbf{f},\ \forall K \in \mathcal{D}(I_0) \setminus \left(\bigcup_{J\in\mathscr{S}(I_0)} \mathcal{D}(J) \right), \tag{4.5}$$

and similarly for $\mathbf{g}$. It follows that

$$\begin{aligned} &\sum_{K\in\mathcal{D}(I_0)\setminus\left(\bigcup_{J\in\mathscr{S}(I_0)}\mathcal{D}(J)\right)} |\vartriangle_K \hat{\mathbf{f}}| \cdot |\vartriangle_K \hat{\mathbf{g}}| \cdot |K| \\ &= \frac{1}{d^2} \left(\sum_{K\in\mathcal{D}(I_0)\setminus\left(\bigcup_{J\in\mathscr{S}(I_0)}\mathcal{D}(J)\right)} |K| \right) |\vartriangle_{I_0} \mathbf{f}| \cdot |\vartriangle_{I_0} \mathbf{g}|. \end{aligned}$$

Therefore, it suffices to verify that

$$\sum_{K\in\mathscr{T}(I_0)} |K| = \frac{1}{d^2} |I_0|, \tag{4.6}$$

where $\mathscr{T}(I_0) := \mathcal{D}(I_0) \setminus \left(\bigcup_{J\in\mathscr{S}(I_0)} \mathcal{D}(J)\right)$. Consider the limiting function $S = \sum_{K\in\mathscr{T}(I_0)} h_K$ (the sum should be understood in both the a.e. on I_0 and $L^2(I_0)$ senses). By the definition (4.2) of the stopping intervals for I_0 we obtain $|S| = d$ a.e. on I_0. In view of orthogonality of Haar functions, it follows that

$$\sum_{K\in\mathscr{T}(I_0)} |K| = \sum_{K\in\mathscr{T}(I_0)} \|h_K\|^2_{L^2(I_0)} = \|S\|^2_{L^2(I_0)} = d^2|I_0|,$$

concluding the proof. □

Remark 4.3. Consider the dyadic Hardy-Littlewood maximal functions $M\mathbf{f}$, $M\widetilde{\mathbf{f}}$ of $\mathbf{f}, \widetilde{\mathbf{f}}$ respectively. We claim that $M\widetilde{\mathbf{f}} \geq (M\mathbf{f}) \circ \Phi$ a.e. on $[0,1)$.

Indeed, note first that $|\widetilde{f}| = |f| \circ \Phi = |f \circ \Phi| = |\widetilde{f}|$, so $|\widetilde{f}|$ is obtained from $|f|$ through the same "small step" transform as $\widetilde{f}$ is obtained through f. It suffices now to note that for all $I \in \mathcal{D}$ and for all $G \in L^\infty(I)$ we have

$$\langle R_I G\rangle_J = \langle G\rangle_{I_\pm},\ \forall J \in \mathscr{S}_\pm(I).$$

4.1.3. Supressing dyadic smoothness constants

We next show that the "small step" construction as given above provides very tight control over dyadic smoothness constants, provided d is large enough.

Lemma 4.4. *Let the weights $w, \widetilde{w}$ be as above. Given $\varepsilon > 0$, assume that $d > (S^d_w - 1)/\varepsilon$. Then, the dyadic smoothness constant $S^d_{\widetilde{w}}$ of the weight $\widetilde{w}$ is less than $1 + \varepsilon$.*

Proof. First of all, it is immediate by rescaling and translation invariance that for all $I \in \mathcal{D}$ and for all weights ρ on I, the dyadic smoothness constant of the weight $(R_I\rho)|_J$ is not larger than S^{d}_ρ, for all $J \in \mathscr{S}(I)$. Therefore, since the transform is given by iteration of the same fundamental transform over $[0,1)$ and all stopping intervals, up to translation and rescaling, it suffices only to verify that

$$\max\left(\frac{\langle\hat{w}\rangle_{K_-}}{\langle\hat{w}\rangle_{K_+}}, \frac{\langle\hat{w}\rangle_{K_+}}{\langle\hat{w}\rangle_{K_-}}\right) \leq 1 + \varepsilon, \qquad \forall K \in \mathcal{D}(I_0) \setminus \left(\bigcup_{J\in\mathscr{S}(I_0)} \mathcal{D}(J)\right), \tag{4.7}$$

where $I_0 = [0,1)$ and $\hat{w} := R_{[0,1)}w$, provided that $d > (S^{\mathrm{d}}_w - 1)/\varepsilon$.

Let $K \in \mathcal{D}(I_0) \setminus \left(\bigcup_{J\in\mathscr{S}(I_0)} \mathcal{D}(J)\right)$ be arbitrary. We have ${}_{\vartriangle K}\hat{w} = (1/d)_{\vartriangle I_0} w$. Moreover, K_+ can be written as a union of stopping intervals (up to a set of zero measure), therefore $\langle\hat{w}\rangle_{K_+} = a\langle w\rangle_{(I_0)_-} + (1-a)\langle w\rangle_{(I_0)_+}$, for some $a \in [0,1]$. It follows that

$$\left|\frac{\langle\hat{w}\rangle_{K_-}}{\langle\hat{w}\rangle_{K_+}} - 1\right| = \frac{|\langle\hat{w}\rangle_{K_-} - \langle\hat{w}\rangle_{K_+}|}{\langle\hat{w}\rangle_{K_+}} \leq \frac{1}{d} \cdot \frac{|\langle w\rangle_{(I_0)_+} - \langle w\rangle_{(I_0)_-}|}{\min(\langle w\rangle_{(I_0)_+}, \langle w\rangle_{(I_0)_-})}.$$

Without loss of generality, we may assume that $\langle w\rangle_{(I_0)_-} \leq \langle w\rangle_{(I_0)_+}$ (the other case is symmetric). Then, we have

$$\frac{1}{d}\cdot\frac{|\langle w\rangle_{(I_0)_+}-\langle w\rangle_{(I_0)_-}|}{\min(\langle w\rangle_{(I_0)_+},\langle w\rangle_{(I_0)_-})}=\frac{1}{d}\cdot\frac{\langle w\rangle_{(I_0)_+}-\langle w\rangle_{(I_0)_-}}{\langle w\rangle_{(I_0)_-}}\leq\frac{1}{d}(S_w^{\mathrm{d}}-1)<\varepsilon.$$

Similarly $\langle\hat{w}\rangle_{K_+}/\langle\hat{w}\rangle_{K_-}<1+\varepsilon$, concluding the proof. □

4.1.4. Respecting dyadic Muckenhoupt characteristics

We next show that the "small step" construction does not ruin dyadic Muckenhoupt constants, up to constants depending only on p. Namely, we claim that $[\widetilde{w},\widetilde{\sigma}]_{A_p,\mathcal{D}}\leq 2^p[w,\sigma]_{A_p,\mathcal{D}}$. To see that, note first that it immediate from translation and rescaling invariance that for all $J\in\mathscr{S}(I_0)$ we have $[\hat{w}|_J,\hat{\sigma}|_J]_{A_p,\mathcal{D}(J)}\leq[w,\sigma]_{A_p,\mathcal{D}(I_0)}$, where $I_0:=[0,1)$ and $\hat{w}:=R_{[0,1)}w$, $\hat{\sigma}:=R_{[0,1)}\sigma$. Therefore, since the transform is given by iteration of the same fundamental transform over $[0,1)$ and all stopping intervals, up to translation and rescaling, it suffices only to verify that

$$\langle\hat{w}\rangle_K\langle\hat{\sigma}\rangle_K^{p-1}\leq 2^p[w,\sigma]_{A_p,\mathcal{D}(I_0)},\qquad\forall K\in\mathcal{D}(I_0)\setminus\left(\bigcup_{J\in\mathscr{S}(I_0)}\mathcal{D}(J)\right).\tag{4.8}$$

Let $K\in\mathcal{D}(I_0)\setminus\left(\bigcup_{J\in\mathscr{S}(I_0)}\mathcal{D}(J)\right)$ be arbitrary. Since K can be written as a union of stopping intervals (up to a set of zero measure), we have $\langle\hat{w}\rangle_K=a\langle w\rangle_{(I_0)_-}+(1-a)\langle w\rangle_{(I_0)_+}$ and $\langle\hat{\sigma}\rangle_K=a\langle\sigma\rangle_{(I_0)_-}+(1-a)\langle\sigma\rangle_{(I_0)_+}$, for some $a\in[0,1]$. Then, the following lemma, whose proof is given in Subsection 8.3 in the Appendix, implies immediately the required result.

Lemma 4.5. *Let $x_1,y_1,x_2,y_2>0$ and $A>0$, such that*

$$x_1y_1^{p-1},\ \left(\frac{x_1+x_2}{2}\right)\left(\frac{y_1+y_2}{2}\right)^{p-1},\ x_2y_2^{p-1}\leq A.$$

Then, there holds

$$(x_1+a(x_2-x_1))(y_1+a(y_2-y_1))^{p-1}\leq 2^pA,\ \forall a\in[0,1].$$

4.1.5. Respecting weighted norms

Finally, we show that weighted norms do not get larger. Consider the function $\widetilde{g}=\widetilde{\mathbf{g}}/\widetilde{w}$. Obviously $\widetilde{g}=g\circ\Phi$. It follows that

$$\|\widetilde{g}\|^{p'}_{L^{p'}(\widetilde{w})}=\int\limits_{[0,1)}|g(\Phi(x))|^{p'}w(\Phi(x))dx=\int\limits_{[0,1)}|g(x)|^{p'}w(x)dx=\|g\|^{p'}_{L^{p}(w)}.\tag{4.9}$$

Similarly $\|\widetilde{f}\|^p_{L^p(\widetilde{\sigma})}=\|f\|^p_{L^p(\sigma)}$, where $\widetilde{f}=\widetilde{\mathbf{f}}/\widetilde{\sigma}$.

4.2. The "small step" construction for Haar shifts

In this section, we describe one variant of the "small step" construction of the previous subsection which exploits the special structure of the martingales in the example of Subsection 3.2

Let $p \in (1, \infty)$ and $M > 2$. Recall the Haar shift T on $[0, 1)$ considered in Subsection 3.2:

$$Tf := 2 \sum_{I \in \mathcal{D}} (\triangle_I f)(h_{I_+} - h_{I_-}).$$

Then, we have

$$\langle Tf, g \rangle = \sum_{I \in \mathcal{D}} |I| (\triangle_I f)(\triangle_{I_+} g - \triangle_{I_-} g), \ \forall f, g \in L^2([0, 1)).$$

Let us first recall the "large step" example of Subsection 3.2. Set $I_0 = [0, 1)$ and $I_n = \left[0, \frac{1}{2^n}\right)$, $J_n = \left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)$, for all $n = 1, 2, \ldots$. Recall that in Subsection 3.2 we showed that there exist bounded weights w, σ on $[0, 1)$ with $\sigma = w^{-1/(p-1)}$,

$$M \leq w([0, 1))\sigma([0, 1))^{p-1}, \qquad [w]_{A_p, \mathcal{D}} \leq 4Me, \qquad w([0, 1)) \sim M, \qquad \sigma([0, 1)) \sim_p 1,$$

with the additional properties $\triangle_I w = \triangle_I \sigma = 0$, for all $I \in \mathcal{D} \setminus \{I_0, I_1, I_2, \ldots\}$, $\triangle_{I_l} w \leq 0$, for all $l = 0, 1, 2, \ldots$, and nonzero bounded functions $f \in L^p(\sigma)$, $g \in L^{p'}(w)$ with

$$\langle f\sigma, T(gw) \rangle \gtrsim_p M \|f\|_{L^p(\sigma)} \|g\|_{L^{p'}(w)}. \tag{4.10}$$

We recall that $g = -\mathbf{1}_{[0,1)}$, so $\mathbf{g} := gw = -w$. Moreover, for the function $\mathbf{f} := f\sigma$ on $[0, 1)$ we have $\langle f \rangle_{[0,1)} = 0$, and for all $I \in \mathcal{D}$ we have $\triangle_I \mathbf{f} \neq 0$ if and only if $I = J_n$ for some positive integer n, in which case $\triangle_I \mathbf{f} > 0$.

Based on this example we want to construct weights $\widetilde{w}, \widetilde{\sigma}$ with $\widetilde{\sigma} = \widetilde{w}^{-1/(p-1)}$, and non-zero functions $\widetilde{f} \in L^p(\widetilde{w})$, $\widetilde{g} \in L^{p'}(\widetilde{\sigma})$ such that (4.10) holds with $\widetilde{f}$, $\widetilde{g}$, $\widetilde{w}$, $\widetilde{\sigma}$ in place of f, g, w, σ. Again, it will be essential that the dyadic smoothness constants of the new weights are as close to 1 as we want. This new example will be used to obtain a "small step" example for the Hilbert transform in Subsection 7.1. For reasons to become apparent there, we will want the martingale differences of the function $\widetilde{\mathbf{g}} := \widetilde{g}\widetilde{w}$ over dyadic intervals of *odd* generation to vanish. Thus, we cannot just mimic naively the "small step" construction of the previous subsection.

4.2.1. "Small step" random walk on a triangle

Consider the $\mathbb{R}^4$-valued martingale X induced by the function $F = (w, \sigma, \mathbf{g}, \mathbf{f})$. Then, we have $\triangle_I X = 0$ for all $I \in \mathcal{D}$ different from $I_0, I_1, I_2, \ldots$ and $J_1, J_2, J_3, \ldots$. Notice that the vectors $\triangle_{I_n} X$, $\triangle_{(I_n)_+} X$ are either linearly independent (in fact orthogonal to

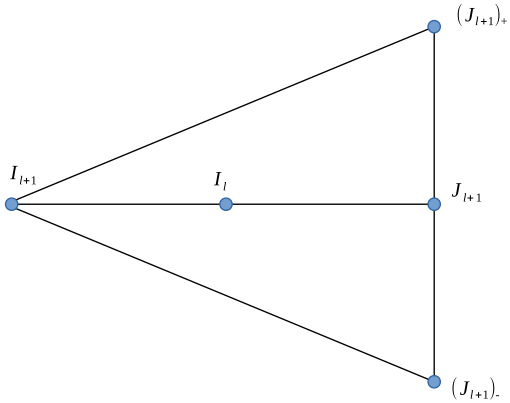


Fig. 4. The triangle corresponding to interval I_l.

each other), or one of them is equal to 0, for all $n = 0, 1, 2, \ldots$. Therefore, the random walk corresponding to the four-dimensional martingale X takes place on the "union" of a family of isosceles triangles in $\mathbb{R}^4$ (maybe degenerate) as in Fig. 4, corresponding to the intervals $I_0, I_1, I_2, \ldots$ respectively.

Starting with the interval $I_0 = [0, 1)$, we replace the constant function $X_0 \equiv \langle X \rangle_{I_0}$ with the function $X^1 = \langle X \rangle_{I_0} + (\vartriangle_{I_0} X)h_{I_0} + (\vartriangle_{(I_0)_+} X)h_{(I_0)_+}$. This function is constant on $(I_0)_- = I_1$ and the children $(I_0)_{+-} = (J_0)_-$, $(I_0)_{++} = (J_0)_+$ of $(I_0)_+ = J_0$. In each of the children of $(I_0)_+$, we just stop, i.e. the function F is constant there, while in the interval $(I_0)_- = I_1$ we repeat this procedure, starting with the constant function $X^1|_{I_1}$, and using the martingale differences of X over I_1, $(I_1)_+$ this time, and then we repeat the same pattern in the interval $(I_1)_- = I_2$, etc. So the random walk corresponding to X consists of rescaled and translated copies of the same pattern, independent from each other. Our main object now is to replace the term $(\vartriangle_{I_n} X)h_{I_n} + (\vartriangle_{(I_n)_+} X)h_{(I_n)_+}$ by a linear combination of Haar functions with "smaller" coefficients, reflecting a "small step" random walk, for all $n = 0, 1, 2, \ldots$.

Choose a sufficiently large positive integer $d > 100$. Condider the model triangle on $\mathbb{R}^2$ with vertices $-e_1$, $e_1 + e_2$ and $e_1 - e_2$, where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Given a dyadic subinterval I of $[0, 1)$ of even generation, we can describe a random walk in I as follows. Starting with the constant function taking value $c_{[0,1)} = 0 \in \mathbb{R}^2$, we replace it with the function $(1/d)h_I e_1 + (1/2d)h_{I_+} e_2$. Notice that the latter function is constant on grandchildren on I. We then repeat the same pattern in the grandchildren of I, and we repeat again this pattern in the grandchildren of the latter intervals, etc. The pattern continues until for some interval J which will have arisen as a grandchild during this process, the current constant value c_J on J is located on the boundary of the triangle. We will say that such intervals J are preliminary stopping intervals. In particular, the preliminary stopping intervals are of even generation. Denote the family of all preliminary stopping intervals by $\widetilde{\mathscr{S}}(I)$.

If J is a preliminary stopping interval such that the constant value c_J on J is located on a side of the model triangle other than its base (that is the vertical side of the triangle), then we replace the constant function c_J on J with the function $c_J + (1/d)h_J e_1 \pm (1/2d)h_J e_2$, where $\pm = +$, respectively $\pm = -$, if c_J is located on the upper, respectively lower, side of the model triangle. Then we repeat this in the grandchildren of J, and then we repeat the pattern in the grandchildren of the latter intervals, etc. The pattern continues until for some interval K which will have arisen as a grandchild during this process, the current constant value c_K on K is located on one of the three vertices of the triangle. We will say that such intervals K are stopping intervals. In particular, these stopping intervals are of even generation.

If J is a preliminary stopping interval such that the constant value c_J on J is located on the base of the triangle, then we replace the constant function c_J on J with the function $c_J + (1/2d)h_J e_2$. Then we repeat this in the grandchildren of J, and then we repeat the pattern in the grandchildren of the latter intervals, etc. The pattern continues until for some interval K which will have arisen as a grandchild during this process, the current constant value c_K on K is located on one of the two vertices of the base. We will also say that such intervals K are stopping intervals. In particular, these stopping intervals are of even generation.

We will denote the family of all stopping intervals by $\mathscr{S}(I)$. We will also denote the family of all stopping intervals J such that c_J is located on the vertex (i.e. $-e_1$) opposite to the base of the model triangle, respectively on the upper vertex (i.e. $e_1 + e_2$) of the base, respectively on the lower vertex (i.e. $e_1 - e_2$) of the base, by $\mathscr{S}_-(I)$, respectively by $\mathscr{S}_{++}(I)$, respectively by $\mathscr{S}_{+-}(I)$. We also set $\mathscr{S}_+(I) = \mathscr{S}_{++}(I) \bigcup \mathscr{S}_{+-}(I)$. We will call the elements of $\mathscr{S}_-(I)$, respectively $\mathscr{S}_+(I)$, left, respectively right, stopping intervals.

Given now a function $G \in L^\infty(I; \mathbb{R}^4)$, the variant of the "small step" transform we are describing here maps it to the function $R_I G := G \circ \psi_I$, where (compare with (4.3))

$$\psi_I(x) = \begin{cases} \psi_{J,I_-}(x), \text{ if } x \in J \text{ for some } J \in \mathscr{S}_-(I) \\ \psi_{J,I_{+\pm}}(x), \text{ if } x \in J \text{ for some } J \in \mathscr{S}_{+\pm}(I) \end{cases}, \forall x \in I. \tag{4.11}$$

The symmetries of the walk imply that $\psi_I : I \to I$ is measure preserving.

The variant of the "small step" transform described here is obtained through iterating the above fundamental transform as follows. We first apply the above construction on the function F, along the interval $[0, 1)$. We thus obtain a function $R_{[0,1)}F \in L^\infty[0, 1); \mathbb{R}^4)$. In each interval in $\mathscr{S}_+(I)$, we just stop (recall that the original function F is constant on the children on I_0), while we apply the above transform on the function $(R_{[0,1)}F)|_I$ along the interval I, for all $I \in \mathscr{S}_-([0, 1))$, and then we stop on every right stopping interval that will have come up, while we repeat the same transform along every left stopping interval that will have come up, etc. Therefore, after this process has been completed we will have obtained a new function $\tilde{F} \in L^\infty([0, 1); \mathbb{R}^4)$.

Recall that the original function F is constant on the children on $(I_n)_+$, for all $n = 0, 1, 2, \ldots$. Note also that $I_{n+1} = (I_n)_-$, for all $n = 0, 1, 2, \ldots$. It follows that $\tilde{F} = F \circ \Phi$,

where $\Phi : [0,1) \to [0,1)$ is the measure-preserving transformation given at almost every point of $[0,1)$ as the composition of all the measure-preserving transformations $\psi_I : I \to I$, where I runs over $[0,1)$ and all left stopping intervals containing that point (note that the order of composition respects inclusion of dyadic intervals). We write $\widetilde{F} = (\widetilde{w}, \widetilde{\sigma}, \widetilde{\mathbf{g}}, \widetilde{\mathbf{f}})$, where tilde denotes just composition with the measure preserving transformation Φ.

Notice that I_0 is an interval of even generation, so its grandchildren are also of even generation, etc. It is then clear that the functions $\widetilde{w}, \widetilde{\sigma}$ are in fact obtained from the functions w, σ respectively thought "small step" transform of order d as in the previous section, but "skipping" intervals of odd generations (i.e. omitting the Haar functions corresponding to them). This means that dyadic intervals I of odd generation "do not split", i.e. $\langle \widetilde{w} \rangle_I = \langle \widetilde{w} \rangle_{I_-} = \langle \widetilde{w} \rangle_{I_+}$, and similarly for $\widetilde{\sigma}$. It is clear that this will be only a minor modification of the construction described in Subsection 4.1. In particular, $\widetilde{w}$, $\widetilde{\sigma}$ are weights on $[0,1)$ with $\widetilde{w}\widetilde{\sigma}^{p-1} = 1$ a.e. on $[0,1)$, and for large enough d the weights $\widetilde{w}$, $\widetilde{\sigma}$ will possess the required dyadic Muckenhoupt characteristic and dyadic smoothness properties.

4.2.2. Getting the damage

We show that the "small step" transform we just described preserves damage for the Haar shift T, i.e. that $\langle \widetilde{\mathbf{f}}, T(\widetilde{\mathbf{g}}) \rangle \gtrsim \langle \mathbf{f}, T(\mathbf{g}) \rangle$.

Lemma 4.6. *Let the functions $\mathbf{f}, \mathbf{g}, \widetilde{\mathbf{f}}, \widetilde{\mathbf{g}}$ be as above. There holds*

$$\sum_{I \in \mathcal{D}} (\vartriangle_I \widetilde{\mathbf{g}})(\vartriangle_{I_+} \widetilde{\mathbf{f}} - \vartriangle_{I_-} \widetilde{\mathbf{f}})|I| \gtrsim \sum_{J \in \mathcal{D}} (\vartriangle_J \mathbf{g})(\vartriangle_{J_+} \mathbf{f} - \vartriangle_{I_-} \mathbf{f})|J|.$$

Proof. First of all, it is immediate by translation and rescaling invariance that

$$\sum_{J \in \mathscr{S}(I_0)} \sum_{K \in \mathcal{D}(J)} (\vartriangle_K \hat{\mathbf{g}})(\vartriangle_{K_+} \hat{\mathbf{f}} - \vartriangle_{K_-} \hat{\mathbf{f}})|K| = \sum_{\substack{J \in \mathcal{D}(I_0) \\ J \neq I_0, (I_0)_+}} (\vartriangle_J \mathbf{g})(\vartriangle_{J_+} \mathbf{f} - \vartriangle_{J_-} \mathbf{f})|J|,$$

where $I_0 = [0,1)$ and $\hat{\mathbf{f}} := R_{[0,1)}\mathbf{f}$, $\hat{\mathbf{g}} := R_{[0,1)}\mathbf{g}$. Note also that $\vartriangle_{(I_0)_+}\mathbf{g} = 0$. Therefore, as in Lemma 4.2, it suffices only to prove the following analog of (4.4):

$$\sum_{I \in \mathscr{T}(I_0)} (\vartriangle_I \hat{\mathbf{g}})(\vartriangle_{I_+} \hat{\mathbf{f}} - \vartriangle_{I_-} \hat{\mathbf{f}})|I| \gtrsim (\vartriangle_{I_0} \mathbf{g})(\vartriangle_{(I_0)_+} \mathbf{f} - \vartriangle_{(I_0)_-} \mathbf{f})|I_0|, \tag{4.12}$$

where $\mathscr{T}(I_0) := \mathcal{D}(I_0) \setminus \left(\bigcup_{J \in \mathscr{S}(I_0)} \mathcal{D}(J) \right)$. Notice that there is an implied absolute constant in the inequality in (4.12), unlike (4.4), where there was just equality. This is no problem (for instance, there will not be accumulation of this constant), since the transform is given by iteration of the same fundamental transform over $[0,1)$ and all left stopping intervals, up to translation and rescaling (essentially, the iterative nature of the transform and translation and rescaling invariance imply that one needs only to verify

the damage inside each triangle separately, and these verifications are independent from each other).

First of all, notice that only intervals in $\widetilde{\mathscr{T}}(I_0) := \mathcal{D}(I_0) \setminus \left(\bigcup_{J\in\widetilde{\mathscr{T}}(I_0)} \mathcal{D}(J)\right)$ that are of even generation may contribute to the sum in (4.12), and for each such interval I we have ${}_{\triangle I}\hat{\mathbf{g}} = (1/d){}_{\triangle I_0}\mathbf{g}$, ${}_{\triangle I_+}\hat{\mathbf{f}} = (1/2d){}_{\triangle(I_0)_+}\mathbf{f}$ and ${}_{\triangle I_-}\hat{\mathbf{f}} = 0 = {}_{\triangle(I_0)_-}\mathbf{f}$. Therefore, it suffices to check that

$$\sum_{I\in\widetilde{\mathscr{T}}_{\mathrm{e}}(I_0)} |I| \gtrsim d^2, \tag{4.13}$$

where $\widetilde{\mathscr{T}}_{\mathrm{e}}(I_0)$ is the family of all intervals in $\widetilde{\mathscr{T}}(I_0)$ that are of even generation. Orthogonality of Haar functions yields

$$\sum_{I\in\widetilde{\mathscr{T}}_{\mathrm{e}}(I_0)} |I| \sim \sum_{I\in\widetilde{\mathscr{T}}_{\mathrm{e}}(I_0)} \left\| h_I e_1 + \frac{1}{2} h_{I_+} e_2 \right\|^2_{L^2(I_0;\mathbb{R}^2)} = \|S\|^2_{L^2(I_0;\mathbb{R}^2)},$$

where we are considering the limiting function $S := \sum_{I\in\widetilde{\mathscr{T}}_{\mathrm{e}}(I_0)} \left(h_I e_1 + \frac{1}{2} h_{I_+} e_2\right)$ (the sum should be understood in both the pointwise a.e. on I_0 and $L^2(I_0;\mathbb{R}^2)$ senses). Rescaling the canonical triangle by d we see that this limiting function is taking values on the boundary of the triangle in $\mathbb{R}^2$ with vertices $(-d,0)$, $(0,d)$, $(0,-d)$. Since the distance of the origin from the boundary of this triangle is $d/\sqrt{5}$, we obtain $|S| \geq d/\sqrt{5}$, therefore $\|S\|^2_{L^2(I_0;\mathbb{R}^2)} \gtrsim d^2$, concluding the proof. □

4.2.3. Respecting weighted norms

Identically to (4.9) we have $\|\widetilde{g}\|^{p'}_{L^{p'}(\widetilde{w})} = \|g\|^{p'}_{L^{p'}(w)}$ and $\|\widetilde{f}\|^{p}_{L^{p}(\widetilde{\sigma})} = \|f\|^{p}_{L^{p}(\sigma)}$, where $\widetilde{g} := \widetilde{\mathbf{g}}/\widetilde{w}$ and $\widetilde{f} := \widetilde{\mathbf{f}}/\widetilde{\sigma}$.

5. Iterated remodeling

In this section we describe the method of iterated remodeling, which is a variant of the powerful method of remodeling, introduced by F. Nazarov in [9].

Throughout this section, for all intervals I, J we denote by $\psi_{I,J}$ the unique orientation-preserving affine transformation mapping I onto J.

5.1. Periodisations

Let $f \in L^\infty([0,1);\mathbb{R}^n)$. For a given interval I and for a given positive integer N, we define the periodisation $\Pi^N_I f$ of f of frequency N over I as the unique periodic function over I of period $\frac{|I|}{2^N}$ consisting of 2^N repeated copies of the function f, i.e. $\Pi^N_I f = f \circ \psi^N_I$, where $\psi^N_I(x) = \psi_{J,I}(x)$ for all $x \in J$, for all $J \in \mathrm{ch}^N(I)$, see Fig. 5 (here we abuse the terminology regarding the use of the term "frequency").

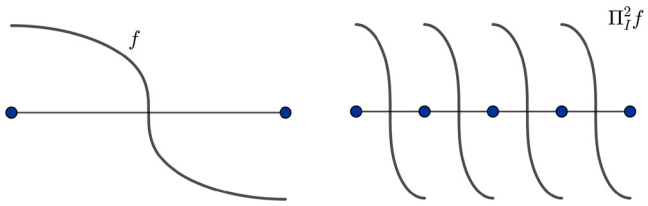


Fig. 5. Periodization $\Pi_I^2 f$ of function f.

Note that $\psi_I^N : I \to I$ is measure preserving. We define the family $\mathscr{E}_N(I)$ of exceptional stopping intervals for I of order N as the family of all intervals in $\mathrm{ch}^N(I)$ that touch the boundary of I (so $\mathscr{E}_N(I)$ has exactly two elements), and the family $\mathscr{R}_N(I)$ of regular stopping intervals for I of order N as the family of all intervals in $\mathrm{ch}^N(I)$ that do not touch the boundary of I.

5.2. From Bourgain's localizing trick to Nazarov remodeling and iterated remodeling

F. Nazarov's method of remodeling [9] had been inspired by a new technique for localizing the action of operators introduced by J. Bourgain in [1]. There, Bourgain showed that UMD property for a Banach space X follows from the boundedness of the Hilbert transform over $L^p(\mathbb{T}; X)$ for all $1 < p < \infty$, where $\mathbb{T}$ denotes the unit circle. Bourgain related estimates for the L^p norm of the Hilbert transform, a non-localized operator, to estimates for the L^p norm of the square function, a well-localized operator, through the trick of iteratively replacing portions of functions with their periodisations.

Bourgain's [1] basic idea was the following. Given a function $f \in L^2(\mathbb{T})$, one can consider its Fourier series

$$\hat{f}(0) + \sum_{m \in \mathbb{Z}\setminus\{0\}} \hat{f}(m) z^m. \tag{5.1}$$

One way to make the action of a bounded in $L^2(\mathbb{T})$ operator on f localized is to create very "large gaps" in expansion (5.1), by considering the function $\widetilde{f}$ with Fourier series

$$\widetilde{f} = \hat{f}(0) + \sum_{m \in \mathbb{Z}\setminus\{0\}} \hat{f}(m) z^{mN_m},$$

where the N_m's are large enough positive integers chosen through an inductive procedure. Here one exploits the fact that z^N converges weakly in (say) $L^2(\mathbb{T})$ to 0 as $N \to \infty$. Note that the "transformed" Fourier series is still a Fourier series.

Given now an X-valued function f (say bounded) on $[0,1)$) (we freely identify $\mathbb{T}$ with $[0,1)$), one can consider its martingale difference decomposition in $L^2([0,1); X)$:

$$f = \langle f \rangle_{[0,1)} + \sum_{I \in \mathcal{D}} \Delta_I f. \tag{5.2}$$

In general, when the Hilbert transform acts on f its action will not be localized, i.e. there will be interactions between martingale differences over different intervals. One could then think of attempting to somehow introduce very large "gaps" in (5.2), inspired from the respective situation in Fourier series. This is not directly possible, and instead one has to notice that the idea in the setting of Fourier series was to replace each z^m with $(z^{N_m})^m$, which is a just a periodisation of z^m. Then one notices that the periodisations of a given martingale difference converge weakly to 0 in (say) L^2 as the frequency increases (see Lemma 7.2). Therefore, one can attempt to replace each martingale difference in (5.2) with a periodisation of it. The frequencies would be chosen large enough through an inductive procedure. Note that the "transformed" martingale difference decomposition should be still a martingale difference decomposition, thus the periodised martingale differences should still somehow respect the hierarchy of dyadic intervals.

Bourgain [1] not only came up with the above intuition, but also found a sleek way to make it precise. Namely, given an X-valued function f (say bounded) on the unit interval $I_0 := [0,1)$, one begins by choosing a frequency $N(I_0)$ and replacing f with its periodisation $\widetilde{f}^1 := \Pi_{I_0}^{N(I_0)} f$. Consider the collection $\mathscr{S}^1 := \mathrm{ch}^{N(I_0)}(I_0)$. Note that

$$\mathbb{E}_{\mathrm{ch}(\mathscr{S}^1)}[\widetilde{f}^1] - \langle f \rangle_{I_0} = \Pi_{I_0}^{N(I_0)}(\Delta_{I_0} f). \tag{5.3}$$

Then, for all $I \in \mathrm{ch}(\mathscr{S}^1)$, one can replace the function $\widetilde{f}^1|_I$ with a periodisation $\Pi_I^{N(I)}(\widetilde{f}^1|_I)$ of it over I, for some choice of frequency $N(I)$. After this has been completed for every interval in $\mathrm{ch}(\mathscr{S}^1)$, one will have obtained a new function $\widetilde{f}^2$ and a new collection of intervals $\mathscr{S}^2 := \bigcup_{I \in \mathrm{ch}(\mathscr{S}^1)} \mathrm{ch}^{N(I)}(I)$. Then, one can repeat this process in each of the intervals in $\mathrm{ch}(\mathscr{S}^2)$ for the function $\widetilde{f}^2$, etc.

One finally obtains a new function $\widetilde{f}$. Note that this function is given as the composition of f with a certain measure-preserving transformation (depending only on the choices of the frequencies), basically because each step in the iterative procedure amounted to composing with a measure-preserving transformation. Notice also that the choices of frequencies of each step of periodisation are separated from each other, so one has really complete freedom in performing them.

It is important to note that the function $\widetilde{f}$ can be obtained as the limit (in any reasonable sense) of a sequence of *averaged* periodisations $\mathbb{E}_{\mathrm{ch}(\mathscr{S}^1)}[\widetilde{f}^1], \mathbb{E}_{\mathrm{ch}(\mathscr{S}^2)}[\widetilde{f}^2], \mathbb{E}_{\mathrm{ch}(\mathscr{S}^3)}[\widetilde{f}^3]$, …, enabling us to keep track of the averages of the new function. It is also essential to note that since the iterative scheme consists in an iteration of the same fundamental construction (that of replacing by a periodisation), up to translating and rescaling, one deduces that an appropriately rescaled and translated version of (5.3) will hold for each iteration over every interval in $\mathrm{ch}(\mathscr{S}^1)$, $\mathrm{ch}(\mathscr{S}^2)$, $\mathrm{ch}(\mathscr{S}^3)$, …, namely

$$\mathbb{E}_{\mathrm{ch}(\mathrm{ch}^{N(I)}(I))}[\widetilde{f}^{k+1}] - \langle \widetilde{f}^k \rangle_I = \Pi_I^{N(I)}(\Delta_I \widetilde{f}^k), \ \forall I \in \mathrm{ch}(\mathscr{S}^k), \ \forall k = 1, 2, \ldots,$$

so each difference $\mathbb{E}_{\mathrm{ch}(\mathscr{S}^{k+1})}[\widetilde{f}^{k+1}] - \mathbb{E}_{\mathrm{ch}(\mathscr{S}^k)}[\widetilde{f}^k]$ can be written as a sum of periodisations of the martingale differences of f over the intervals in $\mathrm{ch}^k([0,1))$. Thus $\widetilde{f}$ satisfies the

original intuition. It is also worth noting that for the purpose of just obtaining estimates it is not necessary to go all the way down to $\widetilde{f}$, one can stop only after a finite number of steps.

J. Bourgain's technique in [1] works really well in the unweighted setting of Banach space valued estimates, but in situations of weighted estimates, such as the setup of Sarason's conjecture, it has the drawback that in general it gives no control over strong dyadic smoothness of weights, even if the original weights are dyadically smooth, basically because it gives no control over averages taken over consecutive dyadic intervals, so it is not well-suited for problems involving fattened A_p characteristics. In order to overcome this difficulty, F. Nazarov [9] came up with the idea of "keeping endpoints", as a means of controlling intervals touching each other.

Namely, one replaces f with the function $\widetilde{f}^1$ which is equal to $\Pi_{I_0}^{N(I_0)} f$ on each interval in $\mathrm{ch}^{N(I_0)}(I_0)$ not touching the boundary of I_0, but equal to just the average $\langle f\rangle_{I_0} = \langle \Pi_{I_0}^{N(I_0)} f\rangle_J$ over each interval $J \in \mathrm{ch}^{N(I_0)}(I_0)$ that touches the boundary of I_0. Moreover, one considers the collection $\mathscr{S}^1$ of intervals in $\mathrm{ch}^{N(I_0)}(I_0)$ that do not touch the boundary of I_0, and simply forgets the ones that touch it.

Then, one follows the same iterative scheme as above, always putting averages over intervals touching the boundary, and then forgetting those intervals. One has again complete freedom in choosing the frequencies, and this allowed F. Nazarov to reduce the estimate of the norm of the Hilbert transform over a *weighted* L^2 space to estimating the norm of the square function over the same weighted L^2 space. Just as before, $\widetilde{f}$ can be realized as the limit of the sequence of averaged counterparts $\mathbb{E}_{\mathrm{ch}(\mathscr{S}^1)}[\widetilde{f}^1]$, $\mathbb{E}_{\mathrm{ch}(\mathscr{S}^2)}[\widetilde{f}^2]$, $\mathbb{E}_{\mathrm{ch}(\mathscr{S}^3)}[\widetilde{f}^3], \ldots$. The latter sequence allowed F. Nazarov to deduce that this process, termed by F. Nazarov remodeling, produces (as will be explained below in 6.1.2) strongly dyadically smooth weights, provided that the original weights are dyadically smooth, precisely because original averages are put in intervals that touch the boundary. Of course, one can again stop only after a finite number of steps.

Although F. Nazarov's remodeling from [9] behaves really well with respect to smoothness, it has the drawback that the new functions are not given just as composition of the original functions with a certain measure-preserving transformation (as was the case in Bourgain's technique [1]) due to putting averages over intervals touching the boundary and then forgetting these intervals. As a consequence, one-weight situations of weights w, σ satisfying $w\sigma^{p-1} = 1$ a.e. on $[0, 1)$, as the ones that we are primarily interested in here, will in general be transformed to two-weight situations of weights $\widetilde{w}, \widetilde{\sigma}$ not satisfying any such relation. To overcome this difficulty and at the same time preserve smoothness, one has essentially to not just forget the intervals that touch the boundary, but rather apply again remodeling in them, and do the same for all intervals touching the boundary that will ever come up. Thus, one can say that one has to apply *iterated* remodeling.

We also note that if one is interested in estimates for the norm of the Hilbert transform over weighted L^p spaces for any $1 < p < \infty$ (not just $p = 2$), then one cannot just reduce

the estimate of this norm to the estimate of the norm of the square function or the Haar multiplier over the same weighted space, but rather one has to use some other slightly more complicated Haar shift, like the one introduced in Subsection 3.2:

$$Tf := 2\sum_{I\in\mathcal{D}} (\triangle_I f)(h_{I_+} - h_{I_-}).$$

This will force us to move one generation deeper during remodeling, that is to consider grandchildren rather than just children of intervals in $\mathscr{S}^1, \mathscr{S}^2, \ldots$, essentially because this Haar shift involves interaction of intervals with their children. We emphasize (and it will become clear in Subsection 6.1) that for the purpose of obtaining examples just for dyadic operators (e.g. Haar multipliers, dyadic maximal function) one can use just children of intervals. The reduction of the estimate for the Hilbert transform to that for the special Haar shift of Subsection 3.2 is done in Subsection 7.1.

5.3. The iterative construction

We now describe in detail iterated remodeling.

Let X be a uniformly bounded $\mathbb{R}^n$-valued martingale on $[0,1)$, induced by a function $F \in L^\infty([0,1);\mathbb{R}^n)$ (one should again think here of the special case of weighted estimates, where $n = 4$ and X is induced by the bounded function $F = (w, \sigma, \mathbf{f}, \mathbf{g})$, where $\mathbf{f} := f\sigma$ and $\mathbf{g} := gw$).

Set $I_0 := [0,1)$ and $\widetilde{F}^0 := F$. Pick a frequency $N(I_0)$ and replace F with the function $\tilde{F}^{I_0} := \Pi_{I_0}^{N(I_0)} F$. We can consider a family $\mathscr{R}_{N(I_0)}(I_0)$ of regular stopping intervals (intervals not touching the boundary) and a family $\mathscr{E}_{N(I_0)}(I_0)$ of exceptional stopping intervals (intervals touching the boundary).

Then, for all $J \in \mathscr{E}_{N(I_0)}(I_0)$, we do the same thing in J for the function $(\tilde{F}^{I_0})|_J = F\circ\psi_{J,I_0}$, with respect to some new choice of frequency $N(J)$, obtaining a family $\mathscr{R}_{N(J)}(J)$ of regular stopping intervals and a family $\mathscr{E}_{N(J)}(J)$ of exceptional stopping intervals. We afterwards repeat this in each new exceptional stopping interval that will have come up, etc. We continue this until the entire I_0 has been covered, up to a Borel set of zero measure, by regular stopping intervals. We note that this will happen because the sum of the measures of the exceptional stopping intervals decays at each step at least geometrically with ratio $1/2$.

After this process has been completed, we will have obtained a new function $\widetilde{F}^1$. We denote by $\mathscr{S}^1$ the family of all regular stopping intervals that will have been collected during this procedure. We also denote by $\hat{\mathscr{S}}^1$ the family of all exceptional stopping intervals that will have been collected during this procedure, together with I_0. We define the starting intervals of order 1 as all elements of the family $\hat{\mathscr{S}}^1$. Note that the elements of $\mathscr{S}^1$ are pairwise disjoint and $\bigcup \mathscr{S}^1 = I_0$ up to a Borel set of zero measure. Note also that $\widetilde{F}^1|_I = F \circ \psi_{I,I_0}$, for all $I \in \mathscr{S}^1$.

For the next step, we repeat the same procedure in the interval I and for the function $\widetilde{F}^1|_I$, for all $I \in \mathrm{ch}^2(\mathscr{S}^1)$ (and not just $\mathrm{ch}(\mathscr{S}^1)$). Here we note that $\widetilde{F}^1|_I = F \circ \psi_{I,J}$

for some *grandchild* J of I_0, for all $I \in \mathrm{ch}^2(\mathscr{S}^1)$. After this has been completed for all intervals in $\mathrm{ch}^2(\mathscr{S}^1)$, we will have obtained a new function $\tilde{F}^2 \in L^\infty([0,1);\mathbb{R}^n)$. We denote by $\mathscr{S}^2$ the family of all regular stopping intervals that will have been collected during this step. Moreover, we denote by $\hat{\mathscr{S}}^2$ the family of all new exceptional stopping intervals that will have been collected during this step, together with all intervals in $\mathrm{ch}^2(\mathscr{S}^1)$. We define the starting intervals of order 2 as all elements of the family $\hat{\mathscr{S}}^2$.

Afterwards, we repeat the same procedure along the interval I and for the function $\tilde{F}^2|_I$, for all $I \in \mathrm{ch}^2(\mathscr{S}^2)$, etc.

After this process has been completed, we will have obtained a sequence of functions $\widetilde{F}^1, \widetilde{F}^2, \widetilde{F}^3, \dots$ and a new function $\widetilde{F} \in L^\infty([0,1);\mathbb{R}^n)$.

5.3.1. Measure-preserving transformation

It is important to note that this process of iterated remodeling amounts just to composition of limiting functions with a certain measure-preserving transformation that depends only on the choice of frequencies. Indeed, is clear that for all $l = 0, 1, 2, \dots$, there exists a measure-preserving transformation $\Psi_l : [0,1) \to [0,1)$ such that $\widetilde{F}^l = \widetilde{F}^{l-1} \circ \Psi_l$, for all $l = 1, 2, \dots$. Then, we have $\widetilde{F} = F \circ \Psi$, where $\Psi : [0,1) \to [0,1)$ is the measure-preserving transformation given at almost every point of $[0,1)$ as the composition of these measure-preserving transformations $\Psi_{\mathbf{1}} \circ \Psi_2 \circ \Psi_3 \circ \dots$. Note that Ψ depends only on the choices of frequecies $N(I),\ I \in \hat{\mathscr{S}} := \bigcup_{k=1}^\infty \hat{\mathscr{S}}^k$.

So in particular, it does not matter whether we apply iterated remodeling with respect to a given choice of frequencies to a martingale as a whole or to each of its coordinates separately with respect to the same choice of frequencies.

5.3.2. Averaged counterparts

Note that the inductive procedure will have also produced the families $\mathscr{S}^1, \mathscr{S}^2, \mathscr{S}^3, \dots$ of all regular stopping intervals that will have been collected during the first, second, third etc respectively step, and the families $\hat{\mathscr{S}}^1, \hat{\mathscr{S}}^2, \hat{\mathscr{S}}^3, \dots$ of all starting intervals of order $1, 2, 3, \dots$ respectively. Then, one can realize $\widetilde{F}$ as the limit, say, pointwise a.e. on $[0,1)$ and in $L^2([0,1);\mathbb{R}^n)$, of the sequence of averaged counterparts $\widetilde{X}^0, \widetilde{X}^1, \widetilde{X}^2, \dots$, where $\widetilde{X}^0 := X_0 \equiv \langle F \rangle_{I_0}$ and $\widetilde{X}^k := \mathbb{E}_{\mathrm{ch}^2(\mathscr{S}^k)}[\widetilde{F}^k]$, for all $k = 1, 2, \dots$.

Remark 5.1. It is clear that for all $k = 1, 2$ we have $\mathbb{E}_{\mathrm{ch}^k(\mathscr{S}^1)}[\widetilde{F}] = \mathbb{E}_{\mathrm{ch}^k(\mathscr{S}^1)}[\widetilde{F}^1] = X_k \circ \Psi$, where recall that $X_k = \mathbb{E}_{\mathrm{ch}^k(I_0)}[F]$. Since the iterative scheme consists in an iteration of the same fundamental construction, up to translating and rescaling, we deduce that

$$\mathbb{E}_{\mathrm{ch}^k(\mathscr{S}^l)}[\widetilde{F}] = X_{2l+k} \circ \Psi, \qquad \forall k = 1, 2, \qquad \forall l = 1, 2, \dots,$$

where $X_{2l+k} = \mathbb{E}_{\mathrm{ch}^{2l+k}(I_0)}[F]$. In particular, the family of all averages of $\widetilde{F}$ over dyadic intervals coincides with the family of all averages of F over dyadic intervals. This had been noted in [9, §10].

5.3.3. Martingale difference decomposition

We now provide a description for the martingale difference decomposition of the function $\widetilde{F}$. Note here that the iterative scheme involved considering *grandchildren* of $\mathscr{S}^1, \mathscr{S}^2, \ldots$, rather than just children. This means that the martingale difference decomposition of $\widetilde{F}$ will involve periodisations of *second order* martingale differences of F, and not just of martingale differences of F (unlike Bourgain's [1] construction and Nazarov's [9] constructions). At the same time, the fact that we do distinguish between intervals that touch the boundary and intervals that do not mean that these periodisations will extend only over intervals that do not touch the boundary, so there will be *quasi-periodisations* rather than just periodisations (like Nazarov's [9] construction, but unlike Bourgain's [1] construction).

Namely, define the second order martingale difference $\Delta_I^2 f$ of a function $f \in L^\infty(I;\mathbb{R}^n)$ over an interval I by

$$\Delta_I^2 f := \mathbb{E}_{\mathrm{ch}^2(I)} f - \langle f\rangle_I \mathbf{1}_I = \Delta_I f + \sum_{J\in \mathrm{ch}(I)} \Delta_J f. \tag{5.4}$$

Moreover, given a frequency N, define the *averaged quasi-periodisation* $\overline{\mathrm{Q\Pi}}_I^N f$ of f of frequency N over I as the function $\overline{\mathrm{Q\Pi}}_I^N f := \mathbb{E}_{\mathscr{E}_N(I)\cup \mathrm{ch}^2(\mathscr{R}_N(I))}[\Pi_I^N f]$, i.e.

$$\overline{\mathrm{Q\Pi}}_I^N f(x) = \begin{cases} (\mathbb{E}_{\mathrm{ch}^2(I)}[f]\circ \psi_{J,I})(x), \text{ if } x \text{ belongs to some } J\in \mathscr{R}_N(I) \\ \\ \langle f\rangle_I, \text{ if } x \text{ belongs to some } J\in \mathscr{E}_N(I) \end{cases}. \tag{5.5}$$

Note that

$$\overline{\mathrm{Q\Pi}}_I^N f - \langle f\rangle_I \mathbf{1}_I = \overline{\mathrm{Q\Pi}}_I^N(\Delta_I^2 f). \tag{5.6}$$

(notice that $\Delta_I^2 f$ is *constant* on the grandchildren of I). It is clear that

$$\widetilde{F}^1(x) = \Pi_J^{N(J)}(F\circ \psi_{J,I_0})(x), \qquad \forall x\in \bigcup \mathscr{R}_{N(J)}(J), \qquad \forall J\in \hat{\mathscr{S}}^1.$$

Therefore, we deduce

$$\widetilde{X}^1 - \widetilde{X}^0 = \sum_{J\in\hat{\mathscr{S}}^1} (\overline{\mathrm{Q\Pi}}_J^{N(J)}(F\circ\psi_{J,I_0}) - \langle F\rangle_{I_0}\mathbf{1}_J), \tag{5.7}$$

which implies

$$\widetilde{X}^1 - \widetilde{X}^0 = \sum_{J\in\hat{\mathscr{S}}^1} \overline{\mathrm{Q\Pi}}_J^{N(J)}(\Delta_J^2(F\circ\psi_{J,I_0})) = \sum_{J\in\hat{\mathscr{S}}^1} \overline{\mathrm{Q\Pi}}_{I_0}^{N(J)}(\Delta_{I_0}^2 F)\circ\psi_{J,I_0}. \tag{5.8}$$

For all $J \in \hat{\mathscr{S}}^1$, we call the function $D_J F := \overline{\mathrm{QII}}_J^{N(J)}(\Delta_J^2(F \circ \psi_{J,I_0}))$ contribution of the starting interval J to the martingale difference decomposition of $\widetilde{F}$.

We emphasize again that the iterative scheme consists in an iteration of the same fundamental construction, up to translating and rescaling. Therefore, an appropriately rescaled and translated copy of (5.8) will hold for each iteration over every interval in the collections $\mathrm{ch}^2(\mathscr{S}^1)$, $\mathrm{ch}^2(\mathscr{S}^2), \ldots$. Therefore, one can write

$$\widetilde{X}^{k+1} - \widetilde{X}^k = \sum_{J \in \hat{\mathscr{S}}^{k+1}} D_J F,$$

where for all $J \in \hat{\mathscr{S}}^{k+1}$ we have $D_J F = \overline{\mathrm{QII}}_J^{N(J)}(\Delta_I^2(F \circ \psi_{J,I}))$ for some $I \in \mathrm{ch}^{2k}(I_0)$, for all $k = 1, 2, \ldots$. The reason for the "$2k$" is again that at the $(k+1)$-th step we repeat the same fundamental process inside each grandchild of each regular stopping interval of the k-th step. In particular

$$\widetilde{F} = \langle F \rangle_{[0,1)} + \sum_{J \in \hat{\mathscr{S}}} D_J F \tag{5.9}$$

in $L^2([0,1); \mathbb{R}^n)$, where $\hat{\mathscr{S}} := \bigcup_{k=1}^{\infty} \hat{\mathscr{S}}^k$ is the family of all starting intervals.

Remark 5.2. Note that $\langle \overline{\mathrm{QII}}_I^N f \rangle_J = \langle f \rangle_I$, for all dyadic subintervals J of I that touch its boundary. In particular $\langle \overline{\mathrm{QII}}_I^N (\Delta_I^2 f) \rangle_J = 0$, for all dyadic subintervals J of I that touch its boundary.

This observation, coupled with (5.8) and a simple inductive argument yields that for all $k = 0, 1, 2 \ldots$, the average of $\widetilde{X}^k$ over every dyadic interval that touches the boundary of $[0,1)$ is equal to $\langle F \rangle_{[0,1)}$. It follows that the average of $\widetilde{F}$ over every dyadic interval that touches the boundary of $[0,1)$ is equal to $\langle F \rangle_{[0,1)}$.

6. The case of dyadic models

In this section we apply iterated remodeling to obtain examples for dyadic models with weights possessing the required smoothness.

6.1. Estimate for Haar multipliers

Let $p \in (1, \infty)$. Let $M > 2$. Let $\delta > 0$ be arbitrarily small. Recall the Haar multiplier T_ε corresponding to any choice of signs $\varepsilon = (\varepsilon_I)_{I \in \mathcal{D}}$:

$$T_\varepsilon f := \sum_{I \in \mathcal{D}} \varepsilon_I (\triangle_I f) h_I.$$

Recall that in Subsection 4.1 we constructed bounded weights $w\sigma$ on $[0,1)$ with $\sigma = w^{-1/(p-1)}$ and

$$M \le [w]_{A_p,\mathcal{D}}, \ \ \langle w\rangle_{[0,1)}\langle\sigma\rangle_{[0,1)}{}^{p-1} \le 2^p 4eM$$

and $S^{\rm d}_w,\ S^{\rm d}_\sigma \le 1+\delta$, and non-zero bounded functions $f \in L^p(\sigma),\ g \in L^{p'}(w)$, such that for the functions $\mathbf{f} = f\sigma,\ \mathbf{g} = gw$ there holds

$$\frac{\sup_\varepsilon |\langle T_\varepsilon(f\sigma), gw\rangle|}{\|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}} = \frac{\sum_{I\in\mathcal{D}} |I|\ \cdot\ |\vartriangle_I \mathbf{f}|\ \cdot\ |\vartriangle_I \mathbf{g}|}{\|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}} \gtrsim_p M. \tag{6.1}$$

We apply the iterated remodeling transform on the martingale induced by the function $(w,\sigma,\mathbf{f},\mathbf{g})$, for an arbitrary choice of frequencies. As it had been observed in 5.3.1, this is the same as applying the iterated remodeling transform separately to each of the functions $w,\sigma,\mathbf{f},\mathbf{g}$, for the same choice of frequencies. Then, the new martingale is induced by the function $(\widetilde{w},\widetilde{\sigma},\widetilde{\mathbf{f}},\widetilde{\mathbf{g}})$, where tilde denotes just composition with the measure preserving-transformation $\Psi : [0,1) \to [0,1)$ of 5.3.1. Then $\widetilde{\sigma},\widetilde{w}$ are weights on $[0,1)$ with $\widetilde{\sigma} = \widetilde{w}^{-1/(p-1)}$ a.e. on $[0,1)$.

6.1.1. Respecting dyadic Muckenhoupt constants

Remark 5.1 shows that for all $I \in \mathcal{D}$ there exists $J \in \mathcal{D}$ (depending only on the choices of frequencies) such that $\langle\widetilde{w}\rangle_I = \langle w\rangle_J$ and $\langle\widetilde{\sigma}\rangle_I = \langle\sigma\rangle_J$. It follows immediately that $[\widetilde{w}]_{A_p,\mathcal{D}} = [w]_{A_p,\mathcal{D}}$.

6.1.2. Dominating strong dyadic smoothness via dyadic one

Let $\varepsilon > 0$. Assume that δ is small enough, so that $(1+\delta)^3 \le 1+\varepsilon$. We claim that $S^{\rm sd}_{\widetilde{w}} \le 1+\varepsilon$. Indeed, let X be the martingale induced by the function w. Recall from 5.3.2 that $\widetilde{w}$ is realized as the limit of the sequence of averaged counterparts $\widetilde{X}^0, \widetilde{X}^1, \widetilde{X}^2, \ldots$. Recall the expression (5.7):

$$\widetilde{X}^1 - \widetilde{X}^0 = \sum_{J\in\hat{\mathscr{S}}^1} (\overline{\rm QII}_J^{N(J)}(w\circ\psi_{J,I_0}) - \langle w\circ\psi_{J,I_0}\rangle_J \mathbf{1}_J).$$

Note that the function $\widetilde{X}^0$ is constant, so $S^{\rm sd}_{\widetilde{X}^0} = 1$, and also that $S^{\rm d}_w \le 1+\delta$ by construction. Then, the following lemma, proved by F. Nazarov in [9, §10], shows that $S^{\rm sd}_{\widetilde{X}^1} \le 1+\varepsilon$. Induction then gives $S^{\rm sd}_{\widetilde{X}^k} \le 1+\varepsilon$, for all $k = 0,1,2,\ldots$. It follows that $S^{\rm sd}_{\widetilde{w}} \le 1+\varepsilon$, independently of the choices of frequencies. The lemma shows that replacing a portion of a strongly dyadically smooth weight with an averaged quasi-periodisation of another dyadically smooth weight preserves the strong dyadic smoothness of the original weight.

Lemma 6.1. *Let w be a weight on an interval $I \in \mathcal{D}$, and assume that $S^{sd}_w \le 1+\varepsilon$ for some $\varepsilon > 0$. Let J be a dyadic subinterval of I, such that w is constant on J. Let v be a weight on J such that $\langle v\rangle_J = \langle w\rangle_J$ and $S^d_v \le 1+\delta$, where $\delta > 0$ satisfies $(1+\delta)^3 \le 1+\varepsilon$. Consider the weight $\widetilde{w} := w + (\overline{QII}_J^N v)\mathbf{1}_J - \langle v\rangle_J \mathbf{1}_J$ on I, i.e.*

$$\widetilde{w}(x) = \begin{cases} w(x), \ \text{if } x \notin J \\ \\ (\overline{Q\Pi}_J^N v)(x), \ \text{if } x \in J \end{cases}, \qquad \forall x \in I.$$

Then, there holds $S^{sd}_{\widetilde{w}} \leq 1 + \varepsilon$.

Proof. Let $K, L \in \mathcal{D}(I)$ be adjacent with $|K| = |L|$. If either both K and L are not contained in J or both K and L touch the boundary of J, we have

$$\frac{\langle \widetilde{w} \rangle_K}{\langle \widetilde{w} \rangle_L} = \frac{\langle w \rangle_K}{\langle w \rangle_L} \leq S^{\mathrm{sd}}_w \leq 1 + \varepsilon.$$

If one of K, L is contained in J and does not touch the boundary of J, then it is clear that $\langle \widetilde{w} \rangle_K = \langle v \rangle_{K'}$ and $\langle \widetilde{w} \rangle_L = \langle v \rangle_{L'}$ for some $K', L' \in \bigcup_{k=0}^2 \mathrm{ch}^k(J)$, therefore

$$\frac{\langle \widetilde{w} \rangle_K}{\langle \widetilde{w} \rangle_L} \leq (S^{\mathrm{d}}_v)^3 \leq (1 + \delta)^3 \leq 1 + \varepsilon,$$

concluding the proof. □

6.1.3. Extending the weights to the entire real line

Consider now the weights $\widetilde{w}', \widetilde{\sigma}'$ on $\mathbb{R}$ given by

$$\widetilde{w}'(x) = \begin{cases} \tilde{w}(x - k), \ \forall x \in (k, k+1), \ \text{if } k \text{ is even} \\ \widetilde{w}(k + 1 - x), \ \forall x \in (k, k+1), \ \text{if } k \text{ is odd} \end{cases}, \qquad \forall k \in \mathbb{Z},$$

and similarly for $\widetilde{\sigma}'$. Obviously $\widetilde{\sigma}' = (\widetilde{w}')^{-1/(p-1)}$. Translation and reflection invariance shows immediately that $[\widetilde{w}']_{A_p, \mathcal{D}} = [\widetilde{w}]_{A_p, \mathcal{D}([0,1))}$. Moreover, translation and reflection invariance yields that $S^{\mathrm{sd}}_{\widetilde{w}'}$ over $[k, k+1)$ is equal to $S^{\mathrm{sd}}_{\widetilde{w}}$, for all $k \in \mathbb{Z}$. Noticing now that for all adjacent $I, J \in \mathcal{D}$ with $|I| = |J|$ whose common endpoint is an integer there holds $\langle \widetilde{w}' \rangle_I = \langle \widetilde{w}' \rangle_J$, we deduce $S^{\mathrm{sd}}_{\widetilde{w}'} = S^{\mathrm{sd}}_{\widetilde{w}}$. Similarly $S^{\mathrm{sd}}_{\widetilde{\sigma}'} = S^{\mathrm{sd}}_{\widetilde{\sigma}}$.

For any $\varepsilon > 0$, one can then achieve $S_{\widetilde{w}'}, S_{\widetilde{\sigma}'} \leq 1 + \varepsilon$ and $[\widetilde{w}']_{A_p} \lesssim_p [\widetilde{w}']_{A_p, \mathcal{D}} = [w]_{A_p, \mathcal{D}([0,1))}$ by taking $\delta > 0$ sufficiently small, per Lemmas 2.1 and 2.2 respectively.

Remark 6.2. We notice that the above estimates yield that the Muckenhoupt characteristic $[\widetilde{w}']_{A_p}$ is comparable to M but in an exponential way with respect to p. In fact, we get $M \leq [\tilde{w}']_{A_p} \leq 2^p 5eM$. If one cares only about dyadic Muckenhoupt characteristics and ignores the "small step" requirement, then as we saw in Section 3 it is possible to give an example with dyadic Muckenhoupt characteristic comparable to M within absolute constants.

6.1.4. Respecting weighted norms

Consider the functions $\widetilde{f}' = (\widetilde{\mathbf{f}}/\widetilde{\sigma})\mathbf{1}_{[0,1)}$, $\widetilde{g}' = (\widetilde{\mathbf{g}}/\widetilde{w})\mathbf{1}_{[0,1)}$ on the real line. Identically to the case of the "small-step" transform, see (4.9), we have $\|\widetilde{f}'\|_{L^p(\widetilde{\sigma}')} = \|f\|_{L^p(\sigma)}$ and $\|\widetilde{g}'\|_{L^{p'}(\widetilde{w}')} = \|g\|_{L^p(w)}$.

6.1.5. Getting the damage

It remains now to verify that we get the desired damage.

Lemma 6.3. *Let* $\mathbf{f}, \mathbf{g}, \widetilde{\mathbf{f}}, \widetilde{\mathbf{g}}$ *be as above. There holds*

$$\sum_{I\in\mathcal{D}} |I| \cdot |\vartriangle_I \widetilde{\mathbf{f}}| \cdot |\vartriangle_I \widetilde{\mathbf{g}}| = \sum_{J\in\mathcal{D}} |J| \cdot |\vartriangle_J \mathbf{f}| \cdot |\vartriangle_J \mathbf{g}|.$$

Proof. First of all, since $\sum_{I\in\mathscr{S}^1} |I| = |I_0|$, where $I_0 := [0,1)$, (5.8) coupled with a translation and rescaling argument yields

$$\sum_{I\in\mathscr{S}^1\cup \mathrm{ch}(\mathscr{S}^1)} |I| \cdot |\vartriangle_I \widetilde{\mathbf{f}}| \cdot |\vartriangle_I \widetilde{\mathbf{g}}| = |I_0| \cdot |\vartriangle_{I_0} \mathbf{f}| \cdot |\vartriangle_{I_0} \mathbf{g}| + \sum_{J\in\mathrm{ch}(I_0)} |J| \cdot |\vartriangle_J \mathbf{f}| \cdot |\vartriangle_J \mathbf{g}|,$$

independently of the choices of frequencies. Since the iterative scheme consists in iteration of the same fundamental construction, up to translating and rescaling, over every interval in $\mathrm{ch}^2(\mathscr{S}^1)$, $\mathrm{ch}^2(\mathscr{S}^2), \ldots$, we deduce

$$\begin{aligned}\sum_{I\in\mathscr{S}^{k+1}\cup \mathrm{ch}(\mathscr{S}^{k+1})} |I| \cdot |\vartriangle_I \widetilde{\mathbf{f}}| \cdot |\vartriangle_I \widetilde{\mathbf{g}}| = &\sum_{J\in\mathrm{ch}^{2k}(I_0)} |J| \cdot |\vartriangle_J \mathbf{f}| \cdot |\vartriangle_J \mathbf{g}| \\ &+ \sum_{J\in\mathrm{ch}(\mathrm{ch}^{2k}(I_0))} |J| \cdot |\vartriangle_J \mathbf{f}| \cdot |\vartriangle_J \mathbf{g}|,\end{aligned}$$

for all $k = 1, 2, \ldots$. This yields immediately the desired result. □

Remark 6.4. Consider the dyadic Hardy-Littlewood maximal functions $M\mathbf{f}, M\widetilde{\mathbf{f}}$ of $\mathbf{f}, \widetilde{\mathbf{f}}$ respectively. Then, similarly to Remark 4.3 we have that the function $|\widetilde{\mathbf{f}}|$ is obtained from the function $|\mathbf{f}|$ through the same iterated remodeling transform as the function $\widetilde{\mathbf{f}}$ is obtained from the function $\mathbf{f}$. Remark 5.1 yields then $M\widetilde{\mathbf{f}} = (M\mathbf{f}) \circ \Psi$ a.e. on $[0,1)$.

This observation, coupled with Remark 4.3, shows that any "large step" family of examples establishing sharpness of weighted estimates for the dyadic Hardy-Littlewood maximal function over $[0,1)$ (see [2]) yields a family of examples (on the entire real line) with weights of arbitrary smoothness achieving that, in exactly the same way that this was done for the Haar multipliers above.

Remark 6.5. We see that in this simple case of dyadic models, the choices of frequencies were irrelevant. It is also clear that one could have considered just children of intervals instead of grandchildren. We will however see that in the more subtle case of the Hilbert

transform, frequencies will have to be chosen appropriately in order to achieve localization of the action of the operator, and considering grandchildren instead of just children will be essential, given the nature of the special Haar shift.

6.2. Muckenhoupt weights taking only two values with prescribed smoothness

We now show how the discussion in Subsection 6.1 implies the result of Proposition 1.5.

Let $p \in (1,\infty)$. Let $Q > 1$. Let $\varepsilon > 0$ be arbitrarily small. Choose $A_0, B_0 > 0$ with $A_0 B_0^{p-1} = Q$. By the results in the appendix we have that there exist $a_1, b_1, a_2, b_2 > 0$, such that $a_1 b_1^{p-1} = a_2 b_2^{p-1} = 1$ and $A_0 = (a_1 + a_2)/2$, $B_0 = (b_1 + b_2)/2$. Consider the weights w, σ on $[0,1)$ given by

$$w := a_1 1_{I_1} + a_2 1_{J_1}, \qquad \sigma := b_1 1_{I_1} + b_2 1_{J_1},$$

where $I_1 = \left[0, \frac{1}{2}\right)$ and $J_1 = \left[\frac{1}{2}, 1\right)$. Then w, σ are bounded, $\sigma = w^{-1/(p-1)}$ and $[w]_{A_p,\mathcal{D}} = w([0,1))\sigma([0,1))^{p-1} = A_0 B_0^{p-1} = Q$. It is also obvious that $S_w^{\mathrm{d}}, S_\sigma^{\mathrm{d}} < \infty$. Choose a sufficiently large positive integer $d > 100$. Apply "small step" transform to the weights w, σ of order d, in order to obtain new weights $\widetilde{w}, \widetilde{\sigma}$ respectively on $[0,1)$, and then the iterated remodeling transform on the functions $\widetilde{w}, \widetilde{\sigma}$, for an arbitrary choice of frequencies (the same for both functions), in order to obtain new weights $\widetilde{w}', \widetilde{\sigma}'$ respectively on $[0,1)$. Extend the latter weights to weights $\widetilde{w}'', \widetilde{\sigma}''$ respectively on $\mathbb{R}$ as in 6.1.3. Then, combining the results of Subsections 4.1 and 6.1 we have $\widetilde{\sigma}'' = \widetilde{w}''^{\,-1/(p-1)}$, $Q \le [\widetilde{w}'']_{A_p} \le 2^p(5/4)Q$ and $S_{\widetilde{w}''}, S_{\widetilde{\sigma}''} \le 1 + \varepsilon$, for small enough ε. Moreover, we have $\widetilde{w}'' \in \{a_1, a_2\}$ a.e. on $\mathbb{R}$, since $\widetilde{w}'$ is obtained from w via composition with measure-preserving transformations.

7. The case of the Hilbert transform

In this section we apply iterated remodeling transform on the martingales in the "small step" example of Subsection 4.2, in order to obtain a "small step" example for the Hilbert transform, proving Theorem 1.1. We then show how this leads to a counterexample to the L^p version of Sarason's conjecture.

7.1. Estimate for the Hilbert transform

We first recall what we achieved in Subsection 4.2. Recall the special Haar shift T from Subsection 4.2:

$$Tf = 2\sum_{I\in\mathcal{D}} (\vartriangle_I f)(h_{I_+} - h_{I_-}).$$

Let $p \in (1,\infty)$ and $M > 2$. Let $\delta > 0$ be arbitrarily small. We constructed bounded weights w, σ on $[0,1)$, such that $\sigma = w^{-1/(p-1)}$,

$$M \le w([0,1))\sigma([0,1))^{p-1},\ [w]_{A_p,\mathcal{D}} \le 2^p 4Me$$

and $w([0,1)) \sim M$, $\sigma([0,1)) \sim_p 1$, and also $S_w^{\mathrm{d}}, S_\sigma^{\mathrm{d}} \le 1+\delta$, and non-zero bounded functions $f \in L^p(\sigma)$, $g \in L^{p'}(w)$, such that

$$\langle f\sigma, T(gw)\rangle = \sum_{J\in\mathcal{D}} (\vartriangle_J(gw))(\vartriangle_{J_+}(f\sigma) - \vartriangle_{J_-}(f\sigma))|J| \ge C_p M \|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}. \quad (7.1)$$

Moreover, by construction for the functions $\mathbf{f} := f\sigma,\ \mathbf{g} =: gw$ there holds $\vartriangle_I \mathbf{g} = 0$, for all dyadic intervals I of *odd* generation, and $(\vartriangle_I \mathbf{g})(\vartriangle_{I_-}\mathbf{f}) = 0 \le (\vartriangle_I \mathbf{g})(\vartriangle_{I_+}\mathbf{f})$, for all dyadic intervals I of *even* generation. Note that then

$$|\langle \mathbf{f}, T(\mathbf{g})\rangle| = \langle \mathbf{f}, T(\mathbf{g})\rangle = \sum_J (\vartriangle_J \mathbf{g})(\vartriangle_{J_+}\mathbf{f})|J|,$$

where the summation runs over all $J \in \mathcal{D}$ that are of *even* generation.

7.1.1. Setting up iterated remodeling

We apply the iterated remodeling transform on the functions $w, \sigma, \mathbf{f}, \mathbf{g}$, for some choices of frequencies to be determined later (the same for all functions), obtaining functions $\widetilde{w}, \widetilde{\sigma}, \widetilde{\mathbf{f}}, \widetilde{\mathbf{g}}$ respectively. We extend $\widetilde{w}, \widetilde{\sigma}$ to weights on the whole real line having the desired smoothness and Muckenhoupt characteristic properties, as in 6.1.3. Let us abuse the notation and denote these extensions by the same letter.

Remark 7.1. From Remark 5.2 we deduce that $\langle \widetilde{w}\rangle_I = w([0,1)) = \widetilde{w}([0,1))$, for all dyadic subintervals I of $[0,1)$ that touch its boundary, and similarly for $\widetilde{\sigma}$. This observation will be crucial later in Subsection 7.2.

We denote by H the Hilbert transform on the real line. We consider the operator $H(\,\cdot\,\widetilde{\sigma})$, acting from $L^p(\widetilde{\sigma})$ into $L^p(\widetilde{w})$. Consider the functions $\widetilde{f} = (\widetilde{\mathbf{f}}/\widetilde{\sigma})\mathbf{1}_{[0,1)}$, $\widetilde{g} = (\widetilde{\mathbf{g}}/\widetilde{w})\mathbf{1}_{[0,1)}$ on the real line. Our goal is to show that if the frequencies are chosen appropriately through an inductive procedure, then one can achieve

$$|\langle \widetilde{\mathbf{f}}, H(\widetilde{\mathbf{g}})\rangle| = |\langle \widetilde{f}\widetilde{\sigma}, H(\widetilde{g}\widetilde{w})\rangle| \gtrsim_p M\|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}. \quad (7.2)$$

Assuming that this has been achieved, we will have (since the Hilbert transform is antisymmetric)

$$\|H\|_{L^p(\widetilde{w})} = \|H(\,\cdot\,\widetilde{\sigma})\|_{L^p(\widetilde{\sigma})\to L^p(\widetilde{w})} \ge \frac{|\langle H(\widetilde{f}\widetilde{\sigma}), \widetilde{g}\widetilde{w}\rangle|}{\|\widetilde{f}\|_{L^p(\widetilde{\sigma})}\|\widetilde{g}\|_{L^{p'}(\widetilde{w})}} = \frac{|\langle \widetilde{f}\widetilde{\sigma}, H(\widetilde{g}\widetilde{w})\rangle|}{\|f\|_{L^p(\sigma)}\|g\|_{L^{p'}(w)}} \gtrsim_p M$$

and hence the desired "small step" example for the Hilbert transform.

7.1.2. Decomposing the bilinear form

We begin by writing the functions $\widetilde{\mathbf{f}}$, $\widetilde{\mathbf{g}}$ as the unconditional sums of their martingale differences in $L^2([0,1))$ (up to a constant) as in (5.9), i.e.

$$\widetilde{\mathbf{f}} = \langle \mathbf{f} \rangle_{[0,1)} + \sum_{I \in \hat{\mathscr{I}}} D_I \mathbf{f}, \qquad \widetilde{\mathbf{g}} = \langle \mathbf{g} \rangle_{[0,1)} + \sum_{I \in \hat{\mathscr{I}}} D_I \mathbf{g}, \tag{7.3}$$

and similarly for $\widetilde{\mathbf{g}}$, where $\hat{\mathscr{I}} := \bigcup_{k=1}^{\infty} \mathscr{I}^k$ is the family of all starting intervals and $D_I\mathbf{f}, D_I\mathbf{g}$ are the contributions of the starting interval I to the martingale differences decomposition of $\widetilde{\mathbf{f}}, \widetilde{\mathbf{g}}$ respectively. Since the Hilbert transform is bounded in $L^2(\mathbb{R})$ and antisymmetric, we have

$$\langle H(\widetilde{\mathbf{g}}\mathbf{1}_{[0,1)}), \widetilde{\mathbf{f}}\mathbf{1}_{[0,1)} \rangle = \sum_{I \in \hat{\mathscr{I}}} \langle H(D_I\mathbf{g}), D_I\mathbf{f} \rangle + \text{cross terms}, \tag{7.4}$$

where the cross terms consist of pairings involving either the average of $\mathbf{f}$ or $\mathbf{g}$ over $[0,1)$ and the contribution of some starting interval, or contributions of different starting intervals.

Our object is to show that the main term in the right-hand side of (7.4) produces the desired damage, while the sum of the cross terms can be forced to be arbitrarily close to 0, through an appropriate choice of frequencies (thus essentially achieving localization of the action of the operator).

7.1.3. Forcing the sum of the cross terms to be arbitrarily small

We need the following lemma, whose statement is mentioned in [9, §12], showing essentially that the functions $D_I\mathbf{f}$, $D_I\mathbf{g}$ oscillate arbitrarily fast for large enough frequency $N(I)$. Recall from (5.8) that for all $I \in \hat{\mathscr{I}}$, there exist mean zero functions ϕ_I, $\psi_I \in L^\infty(I)$ such that $D_I\mathbf{f} = \overline{\mathrm{QII}}_I^{N(I)} \phi_I$ and $D_I\mathbf{g} = \overline{\mathrm{QII}}_I^{N(I)} \psi_I$.

Lemma 7.2. *Let $I \in \mathcal{D}$. Let $\phi \in L^\infty(I)$ with $\langle \phi \rangle_I = 0$. Then, there holds $\overline{Q\Pi}_I^N \phi \to 0$ weakly in $L^q(I)$ as $N \to \infty$, for all $q \in (1,\infty)$ and $\overline{Q\Pi}_I^N \phi \to 0$ weakly* in $L^\infty(I)$ as $N \to \infty$.*

Proof. It is clear from definition (5.5) of averaged quasi-periodisations that for all $N = 3, 4, \ldots$, there holds $\langle h, \overline{\mathrm{QII}}_I^N \phi \rangle = 0$, for all functions h on I that are constant on all intervals in $\mathrm{ch}^N(I)$. Note that $\overline{\mathrm{QII}}_I^N \phi$, $N = 3, 4, \ldots$ are uniformly bounded (say by $\|\phi\|_{L^\infty(I)}$) in $L^\infty(I)$. Then, an "$\frac{\varepsilon}{3}$ argument" yields the desired result. □

Now, for all $I \in \mathcal{D} := \mathcal{D}([0,1))$, set $\mathrm{rk}(I) := -\log_2(\ell(I))$. Recall that $\mathcal{D} = \bigcup_{k=0}^{\infty} \mathcal{D}_k$, where $\mathcal{D}_k := \{I \in \mathcal{D} : \ \mathrm{rk}(I) = k\}$ is finite, for all $k = 0, 1, 2, \ldots$. It follows immediately that one can enumerate the elements of the subset $\hat{\mathscr{I}}$ of $\mathcal{D}$ as $I_0, I_1, I_2, I_3, \ldots$, where $I_0 := [0,1)$, such that for all $0 \leq l < k$ there holds $\mathrm{rk}(I_l) \leq \mathrm{rk}(I_k)$. Note then that in particular, for all $0 \leq l < k$ we have either $I_k \cap I_l = \emptyset$ or $I_k \subset I_l$.

Note also that for all $I \in \hat{\mathscr{I}}$, the functions $\phi_I,\ \psi_I \in L^\infty(I)$ depend only on $\mathbf{f},\ \mathbf{g}$ and the choices of frequencies for starting intervals strictly containing I. Therefore, if for some $k = 1, 2, 3, \dots$ we have already picked $N(I_l)$, for all $l = 0, \dots, k-1$, then by Lemma 7.2 we can choose the frequency $N(I_k)$, in a way depending only on the previous choices and the functions $\mathbf{f},\ \mathbf{g}$, such that

$$T_k := H(\langle \mathbf{f}\rangle_{I_0}\mathbf{1}_{I_0}), D_{I_k}\mathbf{g}\rangle + \langle H(D_{I_k}\mathbf{f}), \langle \mathbf{g}\rangle_{I_0}\mathbf{1}_{I_0})\rangle + \left\langle H(D_{I_k}\mathbf{f}), \sum_{l=0}^{k-1} D_{I_l}\mathbf{g}\right\rangle + \left\langle H\left(\sum_{l=0}^{k-1} D_{I_l}\mathbf{f}\right), D_{I_k}\mathbf{g}\right\rangle$$

is as small in absolute value as we want (since the Hilbert transform is bounded in $L^2(\mathbb{R})$). In particular, we can achieve $|T_k| \le \frac{\varepsilon'}{2^{k+1}}$, where $\varepsilon' := \frac{cC_p}{2} M \|f\|_{L^p(\sigma)} \|g\|_{L^{p'}(w)}$, provided the choice of $N(I_k)$ is allowed to depend also on M, p and the functions w, σ. Here, $c > 0$ is an absolute constant to be determined in Lemma 7.3.

Clearly the sum of cross terms is equal to $\sum_{k=1}^{\infty} T_k$, thus one can force this sum to be less that ε' in absolute value, by choosing the frequencies to be large enough, in a way depending only on M, p and the functions w, σ, f, g.

This way of forcing the sum of the cross terms to be arbitrarily close to 0 in absolute value is essentially the same as in [9, §11]. The choice of ϵ' is also the same as in [9, §11], up to the constant c.

7.1.4. Getting the damage from the main term

We will now show that the main term in the right-hand side of (7.4) produces the desired damage, independently of the above choice of frequencies. More precisely, we will show that

$$\sum_{I \in \hat{\mathscr{I}}^1} \langle H(D_I\mathbf{g}), D_I\mathbf{f}\rangle \ \le -c(\vartriangle_{I_0}\mathbf{g})(\vartriangle_{(I_0)_+}\mathbf{f})|I_0|, \tag{7.5}$$

independently of the choice of frequencies for intervals in $\hat{\mathscr{I}}^1$, where $I_0 := [0, 1)$. Keeping in mind that iterated remodeling as described here moves two generations deep at each step, we deduce through a translation and rescaling argument that

$$\sum_{I \in \hat{\mathscr{I}}} \langle H(D_I\mathbf{g}), D_I\mathbf{f}\rangle \ \le -c \sum_{\substack{J \in \mathcal{D} \\ J \text{ is of even generation}}} (\vartriangle_J\mathbf{g})(\vartriangle_{J_+}\mathbf{f})|J| = -c\langle \mathbf{f}, T(\mathbf{g})\rangle,$$

i.e.

$$-\sum_{I \in \hat{\mathscr{I}}} \langle H(D_I\mathbf{g}), D_I\mathbf{f}\rangle \ge c\langle \mathbf{f}, T(\mathbf{g})\rangle = c|\langle \mathbf{f}, T(\mathbf{g})\rangle|.$$

The last equation, coupled with (7.4), (7.1) and the choice of ε', implies (7.2) (with constant $\frac{cC_p}{2}$), yielding the desired result.

We now establish (7.5). Recall that the regular stopping intervals in $\mathscr{S}^1$ cover I_0 up to a set of zero measure, so it suffices to show that $\langle H(D_I\mathbf{g}), D_I\mathbf{f}\rangle \leq -c(\vartriangle_{I_0}\mathbf{g})(\vartriangle_{(I_0)_+}\mathbf{f})|\bigcup\mathscr{R}_{N(I)}(I)|$, for all $I \in \hat{\mathscr{S}}^1$. Recall from (5.8) that for all $I \in \hat{\mathscr{S}}^1$, $D_I\mathbf{g}$ is just a rescaled and translated copy of $\overline{\mathrm{QII}}_{I_0}^{N(I)}(\Delta_{I_0}^2\mathbf{g})$ over I, and similarly for $\mathbf{f}$. Therefore, it suffices only to prove that

$$\langle H(\overline{\mathrm{QII}}_{I_0}^N(\Delta_{I_0}^2\mathbf{g})), \overline{\mathrm{QII}}_{I_0}^N(\Delta_{I_0}^2\mathbf{f})\rangle \leq -c(\vartriangle_{I_0}\mathbf{g})(\vartriangle_{(I_0)_+}\mathbf{f})|\bigcup\mathscr{R}_N(I_0)|,\ \forall N = 3, 4, \ldots.$$

Let us fix a positive integer $N \geq 3$. Recall that from the definition (5.4) of the second order martingale differences we have

$$\Delta_{I_0}^2\mathbf{g} = (\vartriangle_{I_0}\mathbf{g})h_{I_0} + (\vartriangle_{(I_0)_-}\mathbf{g})h_{(I_0)_-} + (\vartriangle_{(I_0)_+}\mathbf{g})h_{(I_0)_+},$$

and similarly for $\mathbf{f}$. It follows from definition (5.5) of averaged quasi-periodisations, and the facts that $\Delta_{I_0}^2\mathbf{g}$ has mean zero and that it is *constant* on the grandchildren of I_0, that

$$\overline{\mathrm{QII}}_{I_0}^N(\Delta_{I_0}^2\mathbf{g}) = \sum_{J\in\mathcal{G}}[(\vartriangle_{I_0}\mathbf{g})h_J + (\vartriangle_{(I_0)_-}\mathbf{g})h_{J_+} + (\vartriangle_{(I_0)_+}\mathbf{g})h_{J_-}],$$

and similarly for $\mathbf{f}$, where $\mathcal{G} := \mathscr{R}_N(I_0)$. Recall that $\vartriangle_{(I_0)_+}\mathbf{g} = \vartriangle_{(I_0)_-}\mathbf{g} = (\vartriangle_{I_0}\mathbf{g})(\vartriangle_{(I_0)_-}\mathbf{f}) = 0$ and that the Hilbert transform is antisymmetric. It follows that

$$\langle H(\overline{\mathrm{QII}}_{I_0}^N(\Delta_{I_0}^2\mathbf{g})), \overline{\mathrm{QII}}_{I_0}^N(\Delta_{I_0}^2\mathbf{f})\rangle = (\vartriangle_{I_0}\mathbf{g})(\vartriangle_{(I_0)_+}\mathbf{f})\big\langle H\big(\sum_{J\in\mathcal{G}} h_J,\big), \sum_{J\in\mathcal{G}} h_{J_+}\big\rangle.$$

Coupled with the fact that $(\vartriangle_{I_0}\mathbf{g})(\vartriangle_{(I_0)_+}\mathbf{f}) \geq 0$, the following lemma yields then the desired result.

Lemma 7.3. *(a) For all intervals I, J in $\mathbb{R}$ with $|I| = |J|$ and $I \cap J = \emptyset$, there holds*

$$\langle H(h_I), h_{J_+}\rangle + \langle H(h_J), h_{I_+}\rangle < 0.$$

(b) There holds $\big\langle H\big(\sum_{J\in\mathcal{G}} h_J\big), \sum_{J\in\mathcal{G}} h_{J_+},\big\rangle \leq -c\,|\bigcup\mathcal{G}|$, where $c = -\langle H(h_{[0,1)}), h_{[\frac{1}{2},1)}\rangle \in (0,\infty)$.

Proof. (a) First of all, direct computation gives

$$H(h_{[0,1)})(x) = \frac{1}{\pi}\ln\left(\frac{4|x(x-1)|}{(2x-1)^2}\right) \text{ for almost every } x \in \mathbb{R}\setminus\left\{0, \frac{1}{2}, 1\right\},$$

so $H(h_{[0,1)})$ can be identified as a smooth function on $\mathbb{R}\setminus\{0, \frac{1}{2}, 1\}$. Direct computation shows then that $H(h_{[0,1)})$ is strictly increasing and strictly concave in $(1,\infty)$, and strictly decreasing in $\big(\frac{1}{2}, 1\big)$.

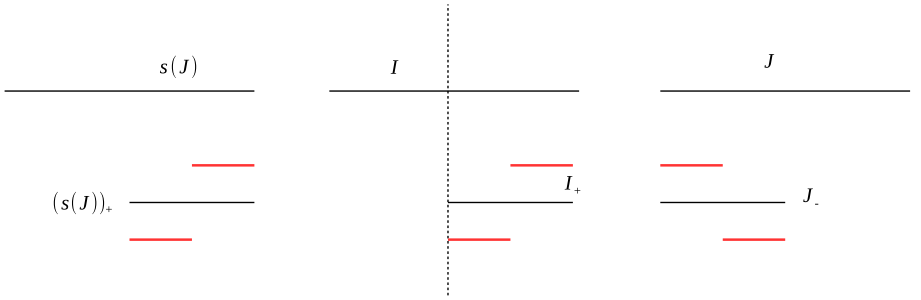


Fig. 6. Illustration of $\langle H(h_J), h_{I_+}\rangle = -\langle H(h_I), h_{J_-}\rangle$.

Let now I, J be intervals in $\mathbb{R}$ with $|I| = |J|$ and $I \cap J = \emptyset$. Without loss of generality, we may assume that $\inf J \geq \sup I$. Note that $H(h_{[0,1)})(1-x) = H(h_{[0,1)})(x)$, for all $x \in \mathbb{R} \setminus \{0, \frac{1}{2}, 1\}$. It follows that for almost every $x \in \mathbb{R}$, if we denote by $s(x)$ the symmetric point to x with respect to the center of I, then we have $H(h_I)(s(x)) = H(h_I)(x)$. Then, a simple symmetry and translation argument, illustrated in Fig. 6, shows that

$$\langle H(h_J), h_{I_+}\rangle = -\langle H(h_I), h_{J_-}\rangle.$$

Therefore, rescaling and translating we obtain

$$\langle H(h_I), h_{J_+}\rangle + \langle H(h_J), h_{I_+}\rangle = \langle H(h_I), h_{J_+} - h_{J_-}\rangle = |I|\langle H(h_{[0,1)}), h_{K_+} - h_{K_-}\rangle,$$

for some interval K in $\mathbb{R}$ with $|K| = 1$ and $\inf K \geq 1$. Therefore, it suffices to prove that the continuous function $\langle H(h_{[0,1)}), h_{[a,a+\frac{1}{2})}\rangle$, $a \in [1, \infty)$ is strictly decreasing. This follows immediately from the fact that the function $H(h_{[0,1)})$ is strictly concave in $(1, \infty)$.

(b) Since $H(h_{[0,1)})$ is strictly decreasing in $\left(\frac{1}{2}, 1\right)$, we have $c := -\langle H(h_{[0,1)}), h_{[\frac{1}{2},1)}\rangle \in (0, \infty)$. Moreover, rescaling and translating we obtain $\langle H(h_I), h_{I_+}\rangle = |I|\langle H(h_{[0,1)}), h_{[\frac{1}{2},1)}\rangle$, for all intervals I in $\mathbb{R}$. Since the intervals in $\mathcal{G}$ are pairwise disjoint and have the same length, we deduce from (a)

$$\begin{aligned}\Big\langle H\Big(\sum_{J\in\mathcal{G}} h_J\Big), \sum_{J\in\mathcal{G}} h_{J_+}\Big\rangle &= \frac{1}{2}\sum_{\substack{J,K\in\mathcal{G}\\ J\neq K}} \left(\langle H(h_J), h_{K_+}\rangle + \langle H(h_K), h_{J_+}\rangle\right) + \sum_{J\in\mathcal{G}}\langle H(h_J), h_{J_+}\rangle \\ &\leq -c\sum_{J\in\mathcal{G}} |J| = -c\Big|\bigcup \mathcal{G}\Big|,\end{aligned}$$

concluding the proof. □

Remark 7.4. The constructions show that for every fixed M, δ and p, one can give examples for the Hilbert transform and Haar multipliers differing only in the function f (and in particular one can take $g = -\mathbf{1}_{[0,1)}$ in both cases).

Remark 7.5. If we were interested just in two-weight estimates, then F. Nazarov's remodeling from [9] would suffice, i.e. one could completely ignore exceptional stopping intervals (except for $[0,1)$ of course), and in fact one could even stop after only a finite number of steps, without losing damage or smoothness of weights. Iteration here only guarantees that the transforms are measure-preserving, so that one-weight situations remain such after applying them.

7.2. Counterexample to L^p version of Sarason's conjecture

Here we describe how the family of examples of Subsection 7.1 will provide through a direct sum of singularities type construction a counterexample to the analog of Sarason's conjecture for every fixed p. Roughly speaking, by direct sum construction one should understand that the unit interval is partitioned into subintervals $J_1, J_2, \ldots$, and then each J_k is equipped with an (appropriately shifted and rescaled) example from the previous section, in such a way that estimates of the norm of the operator blow up as $k \to \infty$.

Fix $p \in (1,\infty)$. Let $\delta > 0$ be sufficiently small. For all $k = 1, 2, \ldots$, by Subsection 7.1 we have that there exist bounded weights w_k, σ_k on $[0,1)$ with

$$[w_k, \sigma_k]_{A_p,\mathcal{D}} \sim_p k, \quad w_k([0,1)) \sim k, \quad \sigma_k([0,1)) \sim_p 1,$$

and $S^{\mathrm{sd}}_{w_k}, S^{\mathrm{sd}}_{\sigma_k} \leq 1+\delta$, and non-zero functions $f_k \in L^p(\sigma_k)$, $g_k \in L^{p'}(w_k)$, such that

$$|\langle H(f_k \sigma_k \mathbf{1}_{[0,1)}), g_k w_k \mathbf{1}_{[0,1)}\rangle| \gtrsim_p k \|f_k\|_{L^p(\sigma_k)} \|g_k\|_{L^{p'}(w_k)}. \tag{7.6}$$

Set $I_0 := [0,1)$ and

$$I_k = \left[0, \frac{1}{2^k}\right), \qquad J_k = \left[\frac{1}{2^k}, \frac{1}{2^{k-1}}\right), \qquad k = 1, 2, \ldots.$$

For all $k = 1, 2, \ldots$, consider the weights $\widetilde{w}_k, \widetilde{\sigma}_k$ on J_k that are obtained as rescaled and shifted copies of the weights $\frac{1}{w_k([0,1))} w_k$, $\frac{1}{\sigma_k([0,1))}\sigma_k$ respectively on the interval $J_k = \left[\frac{1}{2^k}, \frac{2}{2^k}\right)$, i.e.

$$\widetilde{w}_k(x) = \frac{1}{w_k([0,1))} w_k(2^k x - 1), \quad \widetilde{\sigma}_k(x) = \frac{1}{\sigma_k([0,1))}\sigma_k(2^k x - 1), \ \forall x \in J_k,$$

and consider also similarly rescaled and shifted copies $\widetilde{f}_k, \widetilde{g}_k$ of the functions f_k, g_k respectively on the interval J_k. For all $k = 1, 2, \ldots$, we extend the functions $\widetilde{f}_k, \widetilde{g}_k$ on the whole real line by letting them vanish outside of J_k. Consider the weights $\widetilde{w}, \widetilde{\sigma}$ on $[0,1)$ given by $\widetilde{w}(x) = \widetilde{w}_k(x)$, for all $x \in J_k$, for all $k = 1, 2, \ldots$, and similarly for $\widetilde{\sigma}$. We extend the weights $\widetilde{w}, \widetilde{\sigma}$ to weights on the whole real line, as in 6.1.3, and abusing the notation we denote the extended weights by the same letter. Then, translation and rescaling invariance shows that

$$|\langle H(\widetilde{f}_k\widetilde{\sigma}\mathbf{1}_{[0,1)}),\widetilde{g}_k\widetilde{w}\rangle| \gtrsim_p k^{1/p'}\|\widetilde{f}_k\|_{L^p(\widetilde{\sigma})}\|\widetilde{g}_k\|_{L^{p'}(\widetilde{w})}. \tag{7.7}$$

It follows that $\|H(\,\cdot\,\widetilde{\sigma}\mathbf{1}_{[0,1)})\|_{L^p(\widetilde{\sigma})\to L^p(\widetilde{w})}=\infty$. An easy application of the closed graph theorem implies then that there exists $f\in L^p(\widetilde{\sigma})$ with $H(f\widetilde{\sigma}\mathbf{1}_{[0,1)})\notin L^p(\widetilde{w})$. For instance, one can use the facts that

$$\|f\widetilde{\sigma}\mathbf{1}_{[0,1)}\|_{L^1(\mathbb{R})}\le\widetilde{\sigma}([0,1))^{1/p'}\|f\|_{L^p(\widetilde{\sigma})},\ \forall f\in L^p(\widetilde{\sigma}),$$

and that the linear operator $H:L^1(\mathbb{R})\to L^{1,\infty}(\mathbb{R})$ is bounded.

It remains now to prove that the joint "fattened" A_p characteristic of the weights $\widetilde{w},\widetilde{\sigma}$, that is the quantity

$$\sup_{\lambda\in\mathbb{C}_+}\left(\int_{\mathbb{R}}\frac{(\mathrm{Im}(\lambda))^{p-1}}{|x-\lambda|^p}\widetilde{w}(x)dx\right)\left(\int_{\mathbb{R}}\frac{(\mathrm{Im}(\lambda))^{p'-1}}{|x-\lambda|^{p'}}\widetilde{\sigma}(x)dx\right)^{p-1},$$

is finite. As in Subsection 6.1, it suffices to prove that

$$[\widetilde{w},\widetilde{\sigma}]_{A_p,\mathcal{D}([0,1))}\sim_p 1 \tag{7.8}$$

and

$$S^{\mathrm{sd}}_{\widetilde{w}},\ S^{\mathrm{sd}}_{\widetilde{\sigma}}\le 1+\delta. \tag{7.9}$$

Note that translation and rescaling invariance yields immediately that condition (7.8) is fulfilled over J_k, for all $k=1,2,\ldots$. To check it over intervals that are not contained in any J_k, it suffices to note that $\langle\widetilde{w}\rangle_{J_k}=1$, for all $k=1,2,\ldots$, and similarly for $\widetilde{\sigma}$.

Moreover, translation and rescaling invariance yields immediately that condition (7.9) is fulfilled over J_k, for all $k=1,2,\ldots$. Thus, it suffices to check that it still holds for adjacent dyadic intervals of equal length whose common endpoint is also an endpoint of some J_k. To that end, notice that for all $k=1,2,\ldots$, by Remark 7.1 we have $\langle w_k\rangle_{[0,a)}=\langle w_k\rangle_{[a,1)}=w_k([0,1))$, for all $a\in(0,1)$, and similarly for σ_k. It follows that $\langle\widetilde{w}\rangle_J=1$, for all $J\in\mathcal{D}(J_k)$ sharing an endpoint with J_k, and similarly for $\widetilde{\sigma}$, concluding the proof.

Remark 7.6. It is clear that the proof remains valid if we have (7.7) with k raised to any (fixed) positive exponent. Thus, the proof remains valid if we have (7.6) with k raised to any (fixed) exponent greater than $1/p$. Therefore, as long as the Muckenhoupt characteristic estimate in the "large step" examples features an exponent greater than $1/p$, the L^p version of Sarason's conjecture cannot be true.

8. Appendix

8.1. Facts about simply symmetric random walks

We give here the proof of Lemma 4.1. It can be found in any probability theory textbook (see e.g. [8]). We do not follow the notation from Section 2.

Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F} = (\mathcal{F}_n)_{n=0}^{\infty})$ be a filtered probability space. Let $(\omega_n)_{n=1}^{\infty}$ be a sequence of random variables on Ω, such that for all $n = 1, 2, \dots$ the random variable ω_n is $\mathcal{F}_n$-measurable with $\mathbb{P}(\omega_n = 1) = \mathbb{P}(\omega_n = -1) = \frac{1}{2}$, and such that the σ-algebras $\sigma(\omega_n, \omega_{n+1}, \dots)$ and $\mathcal{F}_{n-1}$ are independent. Set $S_0 = 0$ and $S_n = \sum_{k=1}^{n} \omega_k$, for all $n = 1, 2, \dots$. Then, $S = (S_n)_{n=0}^{\infty}$ is a martingale on Ω. In the statement and the proof of the following lemma, we denote $x \wedge y := \min(x, y)$.

Lemma 8.1. *Let $a, b \in (0, \infty)$. Consider the stopping times τ^1, τ^2, τ on Ω given by*

$$\tau^1 := \inf\{n \in \mathbb{N} : \ S_n = b\}, \qquad \tau^2 := \inf\{n \in \mathbb{N} : \ S_n = -a\}, \qquad \tau := \tau^1 \wedge \tau^2.$$

(a) There holds $\tau^1, \tau^2 < \infty$ a.e. on Ω.
(b) There holds $\mathbb{P}(\tau = \tau^1) = \frac{a}{a+b}$ and $\mathbb{P}(\tau = \tau^2) = \frac{b}{a+b}$.

Proof. Let $\theta \in (0, \infty)$ be arbitrary. Consider the martingale M given by

$$M_n := \frac{e^{\theta S_n}}{(\cosh \theta)^n}, \qquad n = 0, 1, 2, \dots$$

(note that $0 < M_n \leq \frac{e^{n\theta}}{(\cosh \theta)^n}$, for all $n = 0, 1, 2, \dots$). By optional sampling theorem, we have that the stopped process $M^{\tau^1} := (M_{n \wedge \tau^1})_{n=0}^{\infty}$ is also a martingale. We notice that

$$0 < M_{n \wedge \tau^1} = \frac{e^{\theta S_{n \wedge \tau^1}}}{(\cosh \theta)^{n \wedge \tau^1}} \leq e^{\theta b}, \qquad \forall n = 0, 1, 2, \dots,$$

thus M^{τ^1} is uniformly bounded. Therefore, by basic convergence facts for martingales it follows that M^{τ^1} is uniformly integrable, therefore there exists $X \in L^1(\Omega)$ such that $M_n^{\tau^1} \to X$ a.e. pointwise on Ω as $n \to \infty$. It is clear that $\lim_{n \to \infty} M_n^{\tau^1}(x) = \frac{e^{\theta S_{\tau^1(x)}(x)}}{(\cosh \theta)^{\tau^1(x)}}$, for all $x \in \Omega$ with $\tau^1(x) < \infty$, and that $M_n^{\tau^1}(x) = \frac{e^{\theta S_n(x)}}{(\cosh \theta)^n}$, for all $n = 0, 1, 2, \dots$, for all $x \in \Omega$ with $\tau^1(x) = \infty$. Then, for all $x \in \Omega$ with $\tau^1(x) = \infty$, we have

$$M_n^{\tau^1}(x) = \frac{e^{\theta S_n(x)}}{(\cosh \theta)^n} \leq \frac{e^{\theta b}}{(\cosh \theta)^n}, \qquad \forall n = 0, 1, 2, \dots,$$

therefore since $\cosh \theta > 1$ we obtain $X(x) = 0$. It follows that

$$\mathbb{E}\left[\frac{e^{\theta S_{\tau^1}}}{(\cosh \theta)^{\tau^1}} \mathbf{1}_{\{\tau^1 < \infty\}}\right] = \mathbb{E}[X] = \mathbb{E}[M_0] = 1,$$

therefore since $S_{\tau^1} = b$ on $\{\tau^1 < \infty\}$ we obtain

$$\mathbb{E}[(\cosh\theta)^{-n}\mathbf{1}_{\{\tau^1<\infty\}}] = e^{-\theta b}.$$

Since $\cosh\theta > 1$, for all $\theta > 0$, taking the limit as $\theta \to 0^+$ and applying the Dominated Convergence Theorem we obtain $\mathbb{P}(\tau^1 < \infty) = 1$. Similarly $\tau^2 < \infty$ a.e. on Ω.

(b) Set $\mathbb{P}(\tau = \tau^1) = p_1$ and $\mathbb{P}(\tau = \tau^2) = p_2$. Then, since $\tau^1, \tau^2 < \infty$ a.e. on Ω we obtain $\tau^1 \neq \tau^2$ a.e. on Ω, therefore $p_1 + p_2 = 1$. We also have $p_1 = \mathbb{P}(S_\tau = b)$ and $p_2 = \mathbb{P}(S_\tau = -a)$. An application of the optional sampling theorem yields $\mathbb{E}[S_\tau] = 0$, i.e. $bp_1 - ap_2 = 0$. Therefore $p_1 = \frac{a}{a+b}$ and $p_1 = \frac{b}{a+b}$. □

8.2. Stopping on the lower hyperbola

We give here the proof of Lemma 3.1.

Let $p \in (1, \infty)$. Let $x, y > 0$ be arbitrary, such that $xy^{p-1} \geq 1$. We claim that there exist $a_1, b_1, a_2, b_2 > 0$ with $a_2 \leq x \leq a_1$ and $b_1 \leq y \leq b_2$, such that $a_1 b_1^{p-1} = a_2 b_2^{p-1} = 1$ and $x = \frac{a_1+a_2}{2}$, $y = \frac{b_1+b_2}{2}$.

Indeed, consider the function $f : (0, 2y) \to (0, \infty)$ given by $f(b) = \frac{1}{b^{p-1}} + \frac{1}{(2y-b)^{p-1}}$, for all $b \in (0, 2y)$. We have $\lim_{b\to 0^+} f(b) = \infty$ and $f(y) = \frac{2}{y^{p-1}} \leq 2x$. Therefore, an application of the Intermediate Value Theorem yields that there exists $b_1 \in (0, y]$ with $f(b_1) = 2x$. Then, we take $b_2 = 2y - b_1$ and $a_1 = b_1^{1-p}$, $a_2 = b_2^{1-p}$.

8.3. Getting a little above the upper hyperbola

We give here the proof of Lemma 4.5.

Let $p \in (1, \infty)$. Let $x_1, y_1, x_2, y_2 > 0$ and $A > 0$, such that

$$x_1 y_1^{p-1}, \ \left(\frac{x_1 + x_2}{2}\right)\left(\frac{y_1 + y_2}{2}\right)^{p-1}, \ x_2 y_2^{p-1} \leq A.$$

We will show that

$$(ax_2 + (1-a)x_1)(ay_2 + (1-a)y_1)^{p-1} \leq 2^p A, \qquad \forall a \in [0, 1].$$

If $x_1 \leq x_2$ and $y_1 \leq y_2$, or $x_1 \geq x_2$ and $y_1 \geq y_2$, then we have nothing to show. Assume now that either $x_2 > x_1$ and $y_1 < y_2$, or $x_1 > x_2$ and $y_2 > y_1$. Replacing if necessary A by $A^{p'-1}$, p by p', and x_i by y_i for $i = 1, 2$, we can without loss of generality assume that there holds $x_1 > x_2$ and $y_2 > y_1$. Set

$$x = \frac{x_1 - x_2}{x_1 + x_2}, \ y = \frac{y_2 - y_1}{y_2 + y_1}, \ B = \frac{A}{\left(\frac{x_1+x_2}{2}\right)\left(\frac{y_1+y_2}{2}\right)^{p-1}}.$$

Then, we have $x, y \in (0, 1)$, $B \geq 1$ and $(1 + x)(1 - y)^{p-1}, \ (1 - x)(1 + y)^{p-1} \leq B$, and we want to show that

$$\sup_{s\in[-1,1]} (1-sx)(1+sy)^{p-1} \le 2^p B.$$

This is clear, because $B \ge 1$ and $(1-sx)(1+sy)^{p-1} \le 2 \cdot 2^{p-1} = 2^p$, for all $s \in [-1,1]$, concluding the proof.

Remark 8.2. Although the above estimate is crude, it can be seen that in general one cannot obtain an estimate better that $2^p/p$ as $p \to \infty$.

8.4. A counterexample

We show here that finiteness of joint Muckenhoupt A_p characteristic does not guarantee two-weight estimates for the Hilbert transform H. We will use a modified version of F. Nazarov's example in [9, p. 1]. Let $p \in (1,\infty)$. Consider the weights w, σ on $\mathbb{R}$ given by

$$w(t) := |t|^{p-1}, \quad \sigma(t) := \begin{cases} |t|^{-p/(p-1)}, & \text{if } |t| > 1 \\ 1, & \text{if } |t| \le 1 \end{cases}, \qquad \forall t \in \mathbb{R}.$$

We show first that $[w,\sigma]_{A_p} < \infty$. It is clear that $\langle w\rangle_{[a,b)}\langle\sigma\rangle_{[a,b)}^{p-1} \lesssim_p 1$, for all $a, b \in \mathbb{R}$ with $-2 \le a < b \le 2$. For all $a \in (1,\infty)$, we have

$$\langle w\rangle_{[0,a)}\langle\sigma\rangle_{[0,a)}^{p-1} \sim_p a^{p-1}\left(\frac{1+1-a^{-1/(p-1)}}{a}\right)^{p-1} \le 2^{p-1}.$$

Then, for all $a, b \in [0,\infty)$ with $0 < a \le \frac{b}{2}$, we have $b - a \ge \frac{b}{2}$, therefore

$$\langle w\rangle_{[a,b)}\langle\sigma\rangle_{[a,b)}^{p-1} \lesssim_p \langle w\rangle_{[0,b)}\langle\sigma\rangle_{[0,b)}^{p-1} \lesssim_p 1.$$

Moreover, for all $a, b \in [0,\infty)$ with $0 < 1 < \frac{b}{2} < a < b$, we have $w(t) \sim_p a^{p-1}$, for all $t \in [a,b)$ and $\sigma(t) \sim_p a^{-p/(p-1)}$, for all $t \in [a,b)$, therefore

$$\langle w\rangle_{[a,b)}\langle\sigma\rangle_{[a,b)}^{p-1} \sim_p a^{p-1}(a^{-p/(p-1)})^{p-1} = a^{-1} < 1.$$

Thus $\langle w\rangle_{[a,b)}\langle\sigma\rangle_{[a,b)}^{p-1} \lesssim_p 1$, for all $a, b \in [0,\infty)$ with $a < b$. This implies $\langle w\rangle_{[-b,-a)}\langle\sigma\rangle_{[-b,-a)}^{p-1} \lesssim_p 1$, for all $a, b \in [0,\infty)$ with $a < b$. Moreover, for all $a, b \in (0,\infty)$, setting $c = \max(a,b)$ and noticing that $b + a \ge c$ we obtain

$$\langle w\rangle_{[-a,b)}\langle\sigma\rangle_{[-a,b)}^{p-1} \lesssim_p \langle w\rangle_{[-c,c)}\langle\sigma\rangle_{[-c,c)}^{p-1} = \langle w\rangle_{[0,c)}\langle\sigma\rangle_{[0,c)}^{p-1} \lesssim_p 1,$$

yielding the desired result.

Consider now the function $f := \mathbf{1}_{[0,1)}$. We have $\|f\|_{L^p(\sigma)} = 1$, just as in [9, p. 1]. Direct computation shows then that $H(f\sigma)(t) = H(\mathbf{1}_{[0,1)})(t) = \frac{1}{\pi}\ln\left(\frac{t}{t-1}\right)$ for almost

every $t \in (1,\infty)$, so since $\lim_{t\to\infty} t \ln\left(\frac{t}{t-1}\right) = 1$ we deduce $H(f\sigma)(t) \sim \frac{1}{t}$ for almost every $t \in (2,\infty)$, thus $|H(f\sigma)(t)|^p w(t) \sim_p \frac{1}{t}$ for almost every $t \in (2,\infty)$, just as in [9, p. 1], thus $H(f\sigma) \notin L^p(w)$.

8.5. Proofs of F. Nazarov's lemmas

We give here the proofs of F. Nazarov's lemmas from [9].

Proof of Lemma 2.1. We follow the proof in [9, §6]. Let $\varepsilon > 0$ be arbitrary. We have $\lim_{\delta\to 0^+}(1+\delta)^{1/\sqrt{\delta}} = \lim_{\delta\to 0^+}(1+\delta) = 1$, therefore there exists $\delta \in \left(0, \frac{1}{4}\right)$ such that

$$(1-2\sqrt{\delta})(1+\delta)^{-2/\sqrt{\delta}} > (1+\varepsilon)^{-1/2}, \quad (1+2\sqrt{\delta})(1+\delta)^{2+2/\sqrt{\delta}} < (1+\varepsilon)^{1/2}.$$

Let now w be a weight on $\mathbb{R}$ with $S_w^{\mathrm{sd}} \le 1+\delta$.

Claim. For all intervals I in $\mathbb{R}$, for all $J \in \mathcal{D}$ with $|J| \le \sqrt{\delta}|I| \le 2|J|$ and containing one of the endpoints of I, there holds $\langle w\rangle_J / \langle w\rangle_I,\ \langle w\rangle_I / \langle w\rangle_J \le (1+\varepsilon)^{1/2}$.

Assume the claim for the moment. Let I be an arbitrary interval in $\mathbb{R}$. There exists $J \in \mathcal{D}$, such that $2|J| \le \sqrt{\delta}|I| \le 4|J|$ and J contains the center of I. Then, by the claim, applied for I_-, J and I_+, J, we have

$$\frac{\langle\rho\rangle_J}{\langle\rho\rangle_{I_-}}, \frac{\langle\rho\rangle_J}{\langle\rho\rangle_{I_+}} \le (1+\varepsilon)^{1/2}, \quad \frac{\langle\rho\rangle_{I_+}}{\langle\rho\rangle_J}, \frac{\langle\rho\rangle_{I_-}}{\langle\rho\rangle_J} \le (1+\varepsilon)^{1/2},$$

therefore $\langle\rho\rangle_{I_+}/\langle\rho\rangle_{I_-},\ \langle\rho\rangle_{I_-}/\langle\rho\rangle_{I_+} \le 1+\varepsilon$, yielding the desired result.

We now prove the claim. Let I be an interval in $\mathbb{R}$, and let $J \in \mathcal{D}$ with $|J| \le \sqrt{\delta}|I| \le 2|J|$, containing one of the endpoints of I.

Set $J_* = \{K \in \mathcal{D} : \ |K| = |J|,\ K \subseteq I\}$ and $I_* = \bigcup J_*$. Clearly $J_* \neq \emptyset$, since $|J| < \frac{1}{2}|I|$. It is clear that

$$\#J_* \le \frac{|I|}{|J|} \le \frac{2}{\sqrt{\delta}}.$$

For all $K \in J_*$, there exist $l \in \{1,\dots,\#J_*\}$ and $J_1,\dots,J_{l+1} \in \mathcal{D}$ of length equal to $|J|$, such that $J_1 = K$, $J_{l+1} = J$ and J_i, J_{i+1} are adjacent or coincide, for all $i = 1,\dots,l$, therefore

$$\frac{\langle w\rangle_K}{\langle w\rangle_J} = \prod_{i=1}^{l} \frac{\langle w\rangle_{J_i}}{\langle w\rangle_{J_{i+1}}} \ge (1+\delta)^{-l} \ge (1+\delta)^{-2/\sqrt{\delta}},$$

thus

$$\frac{\langle w\rangle_{I_*}}{\langle w\rangle_J} = \frac{|J|}{|I_*|} \sum_{K\in J_*} \frac{\langle w\rangle_K}{\langle w\rangle_J} \ge \frac{|J|}{|I_*|}(\#J_*)(1+\delta)^{-2/\sqrt{\delta}} = (1+\delta)^{-2/\sqrt{\delta}}.$$

Note also that $|I_*| \geq |I| - 2|J| \geq (1 - 2\sqrt{\delta})|I|$, therefore

$$\frac{\langle w\rangle_I}{\langle w\rangle_J} \geq \frac{|I_*|}{|I|}\frac{\langle w\rangle_{I_*}}{\langle w\rangle_J} \geq (1-2\sqrt{\delta})(1+\delta)^{-2/\sqrt{\delta}} \geq (1+\varepsilon)^{-1/2}.$$

Set also $J^* = \{K \in \mathcal{D} : \ |K| = |J|, \ K \cap I \neq \emptyset\}$ and $I^* = \bigcup J^*$. It is clear that

$$\#J^* \leq \frac{|I|}{|J|} + 2 \leq \frac{2}{\sqrt{\delta}} + 2.$$

Then, similarly to previously we have

$$\frac{\langle w\rangle_{I^*}}{\langle w\rangle_J} \leq (1+\delta)^{2+2/\sqrt{\delta}}.$$

Note also that $|I_*| \leq |I| + 2|J| \leq (1 + 2\sqrt{\delta})|I|$, therefore

$$\frac{\langle w\rangle_I}{\langle w\rangle_J} \leq \frac{|I^*|}{|I|}\frac{\langle w\rangle_{I^*}}{\langle w\rangle_J} \leq (1+2\sqrt{\delta})(1+\delta)^{2+2/\sqrt{\delta}} \leq (1+\varepsilon)^{1/2},$$

concluding the proof. □

Proof of Lemma 2.2. We follow the proof in [9, §11]. Set $\varepsilon = (25/16)^{1/p} - 1$. Choose $\delta \in \left(0, \frac{1}{4}\right)$ as in the proof of Lemma 2.1 for this ε. Let ρ be a weight on $\mathbb{R}$ with $[\rho]_{A_p,\mathcal{D}} < \infty$ and $S^{\mathrm{sd}}_\rho, S^{\mathrm{sd}}_\tau \leq 1 + \delta$, where $\tau = \rho^{-1/(p-1)}$. Let I be an arbitrary interval in $\mathbb{R}$. By the proof of Lemma 2.1 we have that there exists $J \in \mathcal{D}$ such that

$$\langle \rho\rangle_I \leq (1+\varepsilon)^{1/2}\langle \rho\rangle_J, \quad \langle \tau\rangle_I \leq (1+\varepsilon)^{1/2}\langle \tau\rangle_J,$$

therefore

$$\langle \rho\rangle_I \langle \tau\rangle_I^{p-1} \leq (1+\varepsilon)^{p/2}\langle \rho\rangle_J \langle \tau\rangle_J^{p-1} \leq \frac{5}{4}[\rho]_{A_p,\mathcal{D}}.$$

It follows that $[\rho]_{A_p,\mathcal{D}} \leq [\rho]_{A_p} \leq (5/4)[\rho]_{A_p,\mathcal{D}}$, concluding the proof. □

References

[1] J. Bourgain, Some remarks on Banach spaces in which martingale difference sequences are unconditional, Ark. Mat. 21 (1–2) (1983) 163–168, https://doi.org/10.1007/BF02384306.

[2] S.M. Buckley, Estimates for operator norms on weighted spaces and reverse Jensen inequalities, Trans. Am. Math. Soc. 340 (1) (1993) 253–272.

[3] D.V. Cruz-Uribe, J.M. Martell, C. Pérez, Weights, Extrapolation and the Theory of Rubio de Francia, Operator Theory: Advances and Applications, vol. 215, Birkhäuser/Springer Basel AG, Basel, 2011.

[4] K. Domelevo, P. Ivanisvili, S. Petermichl, S. Treil, A. Volberg, On the failure of lower square function estimates in the non-homogeneous weighted setting, Math. Ann. 374 (2019) 1923–1952.

[5] V.P. Havin, N.K. Nikolski (Eds.), Linear and Complex Analysis. Problem Book 3. Part I, Lecture Notes in Mathematics, vol. 1573, Springer-Verlag, Berlin, 1994.

[6] R. Hunt, B. Muckenhoupt, R.L. Wheeden, Weighted norm inequalities for the conjugate function and Hilbert transform, Trans. Am. Math. Soc. 176 (1973) 227–251.
[7] T.P. Hytönen, The sharp weighted bound for general Calderón–Zygmund operators, Ann. Math. 175 (3) (2012) 1473–1506.
[8] A. Klenke, Probability Theory, A Comprehensive Course, second edition, Springer, 2008.
[9] F. Nazarov, A counterexample to Sarason's conjecture, unpublished manuscript, available at http://users.math.msu.edu/users/fedja/prepr.html.
[10] F. Nazarov, S. Petermichl, S. Treil, A. Volberg, Convex body domination and weighted estimates with matrix weights, Adv. Math. 318 (2017) 279–306.
[11] F. Nazarov, S. Treil, The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis, Algebra Anal. 8 (5) (1996) 32–162; St. Petersburg Math. J. 8 (5) (1997) 721–824.
[12] S. Petermichl, The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical A_p characteristic, Am. J. Math. 129 (5) (2007) 1355–1375.
[13] S. Petermichl, A. Volberg, Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular, Duke Math. J. 112 (2) (2002) 281–305.
[14] E.M. Stein, Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, Princeton, New Jersey, 1993.
[15] S. Treil, A. Volberg, Completely regular multivariate stationary processes and the Muckenhoupt condition, Pac. J. Math. 190 (2) (1999) 361–382.