TWO-WEIGHT INEQUALITY FOR THE HILBERT TRANSFORM: A REAL VARIABLE CHARACTERIZATION, I

MICHAEL T. LACEY, ERIC T. SAWYER,
CHUN-YEN SHEN, and IGNACIO URIARTE-TUERO

Abstract

Let σ and w be locally finite positive Borel measures on $\mathbb{R}$ which do not share a common point mass. Assume that the pair of weights satisfy a Poisson A_2 condition, and satisfy the testing conditions below, for the Hilbert transform H,

$$\int_I H(\sigma \mathbf{1}_I)^2 \, dw \lesssim \sigma(I), \qquad \int_I H(w \mathbf{1}_I)^2 \, d\sigma \lesssim w(I),$$

with constants independent of the choice of interval I. Then $H(\sigma \cdot)$ maps $L^2(\sigma)$ to $L^2(w)$, verifying a conjecture of Nazarov, Treil, and Volberg. The proof has two components, a global-to-local *reduction, carried out in this article, and an analysis of the* local *problem, to be elaborated in a future Part II version of this article.*

1. Introduction

Define a truncated Hilbert transform of a locally bounded signed measure ν by

$$H_{\epsilon,\delta}\nu(x) := \int_{\epsilon<|y-x|<\delta} \frac{d\nu(y)}{y-x}, \quad 0<\epsilon<\delta.$$

Given weights (i.e., locally bounded positive Borel measures) σ and w on the real line $\mathbb{R}$, we consider the following *two-weight norm inequality* for the Hilbert transform,

$$\sup_{0<\epsilon<\delta} \int_{\mathbb{R}} \left|H_{\epsilon,\delta}(f\sigma)\right|^2 dw \le \mathcal{N}^2 \int_{\mathbb{R}} |f|^2 \, d\sigma, \quad f \in L^2(\sigma), \tag{1.1}$$

DUKE MATHEMATICAL JOURNAL
Vol. 163, No. 15, © 2014 DOI 10.1215/00127094-2826690
Received 13 March 2012. Revision received 23 February 2014.
2010 *Mathematics Subject Classification*. Primary 42A50; Secondary 47B38, 42B20.
Lacey's work partially supported by National Science Foundation grant DMS-0968499, Simons Foundation grant 229596, and Australian Research Council grant ARC-DP120100399.
Sawyer's work partially supported by the Natural Sciences and Engineering Research Council of Canada.
Uriarte-Tuero's work partially supported by National Science Foundation grants DMS-0901524 and DMS-1056965, Ministerio de Educación y Ciencias (Spain) grant MTM2010-16232 and MTM2009-14694-C02-01, and by a Sloan Foundation Fellowship.

where $\mathcal{N}$ is the best constant in the inequality, uniform over all truncations of the Hilbert transform kernel. Below, will write the inequality above as $\|H(f\sigma)\|_{L^2(w)} \leq \mathcal{N}\|f\|_{L^2(w)}$, that is, the uniformity over the truncation parameters is suppressed.

The primary question is to find a real variable characterization of this inequality, and the theorem below is an answer to the beautiful conjecture of Nazarov, Treil, and Volberg (see [29, p. 127]). Set

$$P(\sigma, I) := \int_{\mathbb{R}} \frac{|I|}{|I|^2 + \operatorname{dist}(x, I)^2} \sigma(dx),$$

which is approximately the Poisson extension of σ to the upper half-plane, evaluated at $(x_I, |I|)$, where x_I is the center of I.

THEOREM 1.2

Let σ and w be locally finite positive Borel measures on the real line $\mathbb{R}$ with no common point masses. Then, the two-weight inequality (1.1) holds if and only if these three conditions hold uniformly over all intervals I:

$$P(\sigma, I) P(w, I) \leq \mathcal{A}_2, \tag{1.3}$$

$$\int_I \left|H(\mathbf{1}_I \sigma)\right|^2 dw \leq \mathcal{T}^2 \sigma(I), \qquad \int_I \left|H(\mathbf{1}_I w)\right|^2 d\sigma \leq \mathcal{T}^2 w(I). \tag{1.4}$$

We have

$$\mathcal{N} \approx \mathcal{A}_2^{1/2} + \mathcal{T} =: \mathcal{H}, \tag{1.5}$$

where $\mathcal{A}_2$ and $\mathcal{T}$ are the best constants in the inequalities above.

It is well known (see [29]) that the A_2 condition is necessary for the norm inequality, and the inequalities (1.4) are obviously necessary, thus the content of the theorem is the sufficiency of the A_2 and testing inequalities. In the present article we will carry out a *global-to-local* reduction in the proof of sufficiency, while the analysis of the *local* problem will be made in a future article (Part II of this series; see [5]).

The Nazarov–Treil–Volberg conjecture has only been verified before under additional hypotheses on the pair of weights, hypotheses which are not necessary for the two-weight inequality. The so-called *pivotal condition* of [29] is not necessary, as was proved in [9]. The pivotal condition is still an interesting condition: It is all that is needed to characterize the boundedness of the Hilbert transform, together with the maximal function in both directions. But, the boundedness of this triple of operators is decoupled in the two-weight setting (see [24]).

Our argument has these attributes. Certain degeneracies of the pair of weights must be addressed, such as the contribution made by Nazarov, Treil, and Volberg in their innovative 2004 paper [18] (see also [29]), which was further sharpened with the property of *energy* in [9], a crucial property of the Hilbert transform. This theme is further developed in the following pages, with the notion of functional energy explained in Section 5.

The proof should proceed through the analysis of the bilinear form $\langle H(\sigma f), gw\rangle$, as one expects certain paraproducts to appear. Still, the paraproducts have no canonical form, suggesting that the proof is highly nonlinear in f and g. The nonlinear point of view was initiated in [7] and is central to our work here. A particular feature of our arguments is a repeated appeal to certain *quasiorthogonality* arguments, providing (many) simplifications over prior arguments. For instance, we never find ourselves constructing auxiliary measures, and verifying that they are Carleson, a frequent step in many related arguments.

One can phrase a two-weight inequality question for any operator T, a question that became apparent with the foundational paper of Muckenhoupt [12] on A_p weights for the maximal function. Indeed, the case of Hardy's inequality was quickly resolved by Muckenhoupt [11]. The maximal function was resolved by the second author in [26], as well as the fractional integrals, and, essential for this paper, Poisson integrals (see [27]). The latter paper established a result which closely paralleled the contemporaneous $T1$ theorem of David and Journé [1]. This connection, fundamental in nature, was not fully appreciated until the innovative work of Nazarov, Treil, and Volberg [14]–[16] in developing a nonhomogeneous theory of singular integrals. The two-weight problem for dyadic singular integrals was only resolved recently in [17]. Partial information about the two-weight problem for singular integrals in [21] was basic to the resolution of the A_2 conjecture in [3], and several related results (see [4], [6], [21], [22]). Our result is the first real variable characterization of a two-weight inequality for a continuous singular integral.

Interest in the two-weight problem for the Hilbert transform arises from its natural occurrence in questions related to operator theory (see [20], [25]), spectral theory (see [20]), model spaces (see [23]), and analytic function spaces (see [10]). In the context of operator theory, Sarason posed the conjecture (see [2]) that the Hilbert transform would be bounded if the pair of weights satisfied the (full) Poisson A_2 condition. This was disproved by Nazarov [13]. Advances on these questions have been linked to a finer understanding of the two-weight question (see, e.g., [19], [20]) built upon Nazarov's counterexample.

2. Dyadic grids and Haar functions

2.1. Choice of truncation

We have stated the main theorem with *hard* cutoffs in the truncation of the Hilbert transform. There are many possible variants in the choice of truncation; moreover, the proof of sufficiency requires a different choice of truncation.

Consider a truncation given by

$$\tilde{H}_{\alpha,\beta}(\sigma f)(x) := \int f(y) K_{\alpha,\beta}(y-x)\sigma(dy),$$

where $K_{\alpha,\beta}(y)$ is chosen to minimize the technicalities associated with off-diagonal considerations. Specifically, set $K_{\alpha,\beta}(0) = 0$, and otherwise $K_{\alpha,\beta}(y)$ is odd and for $y > 0$,

$$K_{\alpha,\beta}(y) := \begin{cases} -\frac{y}{\alpha^2} + \frac{2}{\alpha} & 0 < y < \alpha, \\ \frac{1}{y} & \alpha \le y \le \beta, \\ -\frac{y}{\beta^2} + \frac{2}{\beta} & \beta < y < 2\beta, \\ 0 & 2\beta \le y. \end{cases}$$

This is a C^1 function on $(0, 2\beta)$, and is Lipschitz, convex, and monotone on $(0, \infty)$.

We now argue that we can use these truncations in the proof of the sufficiency bound of our main theorem.

PROPOSITION 2.1

If the pair of weights σ, w satisfy the A_2 bound (1.3), then one has the uniform norm estimate with the hard *truncations (1.1) if and only if one has uniform norm estimate for the* smooth *truncations,*

$$\sup_{0<\alpha<\beta} \left\| \tilde{H}_{\alpha,\beta}(\sigma f) \right\|_w \le \mathcal{N} \| f \|_\sigma.$$

Indeed, $|H_{\alpha,\beta}(\sigma f) - \tilde{H}_{\alpha,\beta}(\sigma f)| \lesssim A_\alpha(\sigma |f|) + A_\beta(\sigma |f|)$, where these last two operators are *single-scale* averages, namely,

$$A_\alpha(\sigma\phi)(x) = \alpha^{-1} \int_{(x-3\alpha, x+3\alpha)} \phi(y)\sigma(dy).$$

But, the (simple) A_2 bound is all that is needed to provide a uniform bound on the operators $A_\alpha(\sigma\phi)$. So the proposition follows.

Henceforth we use the truncations $\tilde{H}_{\alpha,\beta}$, and we suppress the tilde in the notation. The particular choice of truncation is motivated by this off-diagonal estimate on the kernels.

PROPOSITION 2.2
Suppose that $2|x-x'|<|x-y|$; *then*

$$K_{\alpha,\beta}(y-x')-K_{\alpha,\beta}(y-x)=C_{x,x',y}\frac{x'-x}{(y-x)(y-x')},$$
$$\text{where } C_{x,x',y}=1, 2\alpha<|x-y|<\frac{1}{2}\beta, \tag{2.3}$$

and is otherwise positive and never more than 4.

Proof
The assumptions imply that $y-x'$ and $y-x$ have the same sign. Assume without loss of generality that $0<y-x'<y-x$. If $2\alpha<|x-y|<\beta/2$, it follows that $\alpha<|x'-y|<\beta$, and so by the definition,

$$K_{\alpha,\beta}(y-x')-K_{\alpha,\beta}(y-x)=\frac{1}{y-x'}-\frac{1}{y-x}=\frac{x'-x}{(y-x)(y-x')}.$$

And, in the general case, we have $|\frac{d}{dt}K_{\alpha,\beta}(t)|\le 4t^{-2}$, so that

$$0\le K_{\alpha,\beta}(y-x')-K_{\alpha,\beta}(y-x)\le\int_{y-x'}^{y-x}\frac{4}{t^2}\,dt=4\frac{x'-x}{(y-x)(y-x')}. \qquad \square$$

2.2. Dyadic grids
A collection of intervals $\mathcal{G}$ is a *grid* if for all $G,G'\in\mathcal{G}$, we have $G\cap G'\in\{\emptyset,G,G'\}$. By a *dyadic grid* we mean a grid $\mathcal{D}$ of intervals of $\mathbb{R}$ such that for each interval $I\in\mathcal{D}$, the subcollection $\{I'\in\mathcal{D}:|I'|=|I|\}$ partitions $\mathbb{R}$, aside from endpoints of the intervals. In addition, the left and right halves of I, denoted by $I_\pm$, are also in $\mathcal{D}$.

For $I\in\mathcal{D}$, the left and right halves $I_\pm$ are referred to as the *children* of I. We denote by $\pi_{\mathcal{D}}I$ the unique interval in $\mathcal{D}$ having I as a child, and we refer to $\pi_{\mathcal{D}}I$ as the $\mathcal{D}$*-parent* of I.

We will work with subsets $\mathcal{F}\subset\mathcal{D}$. We say that I has $\mathcal{F}$*-parent* $\pi_{\mathcal{F}}I=F$ if $F\in\mathcal{F}$ is the minimal element of $\mathcal{F}$ that contains I.

2.3. Haar functions
Let σ be a weight on $\mathbb{R}$, one that does not assign positive mass to any endpoint of a dyadic grid $\mathcal{D}$. If $I\in\mathcal{D}$ is such that σ assigns nonzero weight to both children of I, the associated Haar function is

$$h_I^\sigma:=\sqrt{\frac{\sigma(I_-)\sigma(I_+)}{\sigma(I)}}\Big(-\frac{I_-}{\sigma(I_-)}+\frac{I_+}{\sigma(I_+)}\Big).$$

In this definition, we are identifying an interval with its indicator function, and we will do so throughout the remainder of the paper. This is an $L^2(\sigma)$-normalized function, and has σ-integral zero. For any dyadic interval I_0, we have that $\{\sigma(I_0)^{-1/2} I_0\} \cup \{h_I^\sigma : I \in \mathcal{D}, I \subset I_0\}$ is an orthogonal basis for $L^2(I_0, \sigma)$.

We will use the notation $\hat{f}(I) = \langle f, h_I^\sigma \rangle_\sigma$, as well as

$$\Delta_I^\sigma f = \langle f, h_I^\sigma \rangle_\sigma h_I^\sigma = I_+ \mathbb{E}_{I_+}^\sigma f + I_- \mathbb{E}_{I_-}^\sigma f - I \mathbb{E}_I^\sigma f.$$

The second equality is the familiar martingale difference equality, and so we will refer to $\Delta_I^\sigma f$ as a *martingale difference*. It implies the familiar telescoping identity $\mathbb{E}_J^\sigma f = \sum_{I : I \supsetneq J} \mathbb{E}_J^\sigma \Delta_I^\sigma f$.

For any function the *Haar support* of f is the collection $\{I \in \mathcal{D} : \hat{f}(I) \neq 0\}$.

2.4. *Good-bad decomposition*

With a choice of dyadic grid $\mathcal{D}$ understood, we say that $J \in \mathcal{D}$ is (ϵ, r)-*good* if and only if for all intervals $I \in \mathcal{D}$ with $|I| \geq 2^{r-1}|J|$, the distance from J to the boundary of *either child* of I is at least $|J|^\epsilon |I|^{1-\epsilon}$.

For $f \in L^2(\sigma)$, we set $P_{\mathrm{good}}^\sigma f = \sum_{\substack{I \in \mathcal{D} \\ I \text{ is } (\epsilon, r)\text{-good}}} \Delta_I^\sigma f$. The projection $P_{\mathrm{good}}^w g$ is defined similarly. To make the two reductions below, one must make a *random* selection of grids, as is detailed in [9] and [29]. The use of random dyadic grids has been a basic tool since the foundational work of [14]–[16]. Important elements of the suppressed construction of random grids are the following.

(1) It suffices to consider a single dyadic grid $\mathcal{D}$.

(2) For any fixed $0 < \epsilon < 1/2$, we can choose an integer r sufficiently large so that it suffices to consider f such that $f = P_{\mathrm{good}}^\sigma f$, and likewise for $g \in L^2(w)$. Namely, it suffices to estimate the constant below, for an arbitrary dyadic grid $\mathcal{D}$,

$$\left| \langle H_\sigma f, g \rangle_w \right| \leq \mathcal{N}_{\mathrm{good}} \| f \|_\sigma \| g \|_w,$$

where it is required that $f = P_{\mathrm{good}}^\sigma \in L^2(\sigma)$ and $g = P_{\mathrm{good}}^w \in L^2(w)$.

That the functions are good is, at some moments, an essential property. We suppress it in notation, however, taking care to emphasize in the text those places in which we appeal to the property of being good.

A reduction, using randomized dyadic grids, allows one the extraordinarily useful reduction in the next lemma. This is a well-known reduction, due to Nazarov, Treil, and Volberg, explained in full detail in the current setting in [18, Section 4]. Below, $\mathcal{H}$ is as in (1.5), the normalized sum of the A_2 and testing constants.

LEMMA 2.4

For all sufficiently small ϵ, and sufficiently large r, this holds. Suppose that for any *dyadic grid $\mathcal{D}$, such that no endpoint of an interval $I \in \mathcal{D}$ is a point mass for σ or w,*[†] *we have*

$$\left|\langle H_\sigma P^\sigma_{\text{good}} f, P^w_{\text{good}} g\rangle_w\right| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w. \tag{2.5}$$

Then, the same inequality holds without the projections P^σ_{good}, and P^w_{good}.

Inequality (2.5) should be understood as an inequality, uniform over the class of smooth truncations of the Hilbert transform. But, we can suppress this in the notation without causing confusion. The bilinear form only needs to be controlled for (ϵ, r)-*good functions f and g, goodness being defined with respect to a fixed dyadic grid.* Suppressing the notation, we write *good* for (ϵ, r)-*good*, and it is always assumed that the dyadic grid $\mathcal{D}$ is fixed, and only good intervals are in the Haar support of f and g, though is also suppressed in the notation.

3. The global-to-local reduction

The goal of this section is to reduce the analysis of the bilinear form in (2.5) to the local estimate, (3.4). It is sufficient to assume that f and g are supported on an interval I^0; by trivial use of the interval testing condition, we can further assume that f and g are of integral zero in their respective spaces. Thus, f is in the linear span of (good) Haar functions h^σ_I for $I \subset I^0$, and similarly for g, and

$$\langle H_\sigma f, g\rangle_w = \sum_{I,J : I,J \subset I^0} \langle H_\sigma \Delta^\sigma_I f, \Delta^w_J g\rangle_w.$$

The argument is independent of the choice of truncation that implicitly appears in the inner product above.

The double sum is broken into different summands. Many of the resulting cases are elementary, and we summarize these estimates as follows. Define the bilinear form

$$B^{\text{above}}(f, g) := \sum_{I : I \subset I^0} \sum_{J : J \Subset I_J} \mathbb{E}^\sigma_{I_J} \Delta^\sigma_I f \cdot \langle H_\sigma I_J, \Delta^w_J g\rangle_w,$$

where here and throughout, $J \Subset I$ means $J \subset I$ and $2^r|J| \le |I|$. In addition, the argument of the Hilbert transform, I_J, is the child of I that contains J, so that $\Delta^\sigma_I f$ is constant on I_J. Define $B^{\text{below}}(f, g)$ in the dual fashion.

[†]This set of dyadic grids that fail this condition have probability zero in standard constructions of the random dyadic grids.

LEMMA 3.1
We have, with the notation of (1.5),

$$\left|\langle H_\sigma f, g\rangle_w - B^{\text{above}}(f,g) - B^{\text{below}}(f,g)\right| \lesssim \mathcal{H}\|f\|_\sigma\|g\|_w.$$

This is a common reduction in a proof of a $T1$ theorem, and in the current context, it only requires goodness of intervals and the A_2 condition. For a proof, one can consult [29] and [18]. The lemma is specifically phrased and proved in this way in [7, Section 8].

These definitions are needed to phrase the global-to-local reduction. The following definition depends upon the essential energy inequality (4.7) in Section 4.

Definition 3.2
Given any interval F_0, define $\mathcal{F}_{\text{energy}}(F_0)$ to be the maximal subintervals $F \subsetneq F_0$ such that

$$P(\sigma F_0, F)^2 E(w,F)^2 w(F) > 10C_0\mathcal{H}^2\sigma(F),$$

where $E(w,F)$ is defined in (4.6) and where C_0 is the constant in Proposition 4.8. We have $\sigma(\bigcup\{F : F \in \mathcal{F}(F_0)\}) \le (\sigma/10)(F_0)$.

Definition 3.3
Let I_0 be an interval, and let $\mathcal{S}$ be a collection of disjoint intervals contained in I_0. A function $f \in L^2_0(I_0,\sigma)$ is said to be *uniform (with respect to $\mathcal{S}$)* if the following conditions are met.
(1) Each energy stopping interval $F \in \mathcal{F}_{\text{energy}}(I_0)$ is contained in some $S \in \mathcal{S}$.
(2) The function f is constant on each interval $S \in \mathcal{S}$.
(3) For any interval $I \subset I_0$ which is not contained in any $S \in \mathcal{S}$, $\mathbb{E}^\sigma_I|f| \le 1$.
We will say that g is *weakly adapted* to a function f uniform with respect to $\mathcal{S}$ if $J \Subset S$ for some interval $S \in \mathcal{S}$ implies that $\langle g, h^w_J\rangle_w = 0$. We will also say that g is *weakly adapted to $\mathcal{S}$*.

The constant $\mathcal{L}$ is defined as the best constant in the *local estimate*

$$\left|B^{\text{above}}(f,g)\right| \le \mathcal{L}\left\{\sigma(I_0)^{1/2} + \|f\|_\sigma\right\}\|g\|_w, \tag{3.4}$$

where f, g are of mean zero on their respective spaces, supported on an interval I_0. Moreover, f is uniform and g is weakly adapted to f. The inequality above is homogeneous in g, but not f, since the term $\sigma(I_0)^{1/2}$ is motivated by the bounded averages property of f.

THEOREM 3.5 (Global-to-local reduction)
We have

$$\left|B^{\text{above}}(f,g)\right| \lesssim \{\mathscr{H} + \mathscr{L}\}\|f\|_\sigma \|g\|_w .$$

The same inequality holds for the dual form $B^{\text{below}}(f,g)$.

A reduction of this type is a familiar aspect of many proofs of a $T1$ theorem, proved by exploiting standard off-diagonal estimates for Calderón–Zygmund kernels, but in the current setting, it is a much deeper fact, a consequence of the *functional energy inequality* of Section 5. We make the following construction for an $f \in L^2(I^0,\sigma)$, of σ-integral zero. Add I^0 to $\mathcal{F}$, and set $\alpha_f(I^0) := \mathbb{E}^\sigma_{I^0}|f|$. In the inductive stage, if $F \in \mathcal{F}$ is minimal, add to $\mathcal{F}$ those maximal descendants F' of F such that $F' \in \mathcal{F}_{\text{energy}}(F)$ or $\mathbb{E}^\sigma_{F'}|f| \geq 10\alpha_f(F)$. Then define

$$\alpha_f(F') := \begin{cases} \alpha_f(F) & \mathbb{E}^\sigma_{F'}|f| < 2\alpha_f(F), \\ \mathbb{E}^\sigma_{F'}|f| & \text{otherwise.} \end{cases}$$

If there are no such intervals F', the construction stops. We refer to $\mathcal{F}$ and $\alpha_f(\cdot)$ as *Calderón–Zygmund stopping data for f*, following the terminology of [7, Definition 3.5]. Their key properties are collected here.

LEMMA 3.6
For $\mathcal{F}$ and $\alpha_f(\cdot)$ as defined above, the following hold.

(1) *The dyadic interval I_0 is the maximal element of $\mathcal{F}$.*

(2) *For all $I \in \mathcal{D}$, $I \subset I^0$, we have $\mathbb{E}^\sigma_I|f| \leq 10\alpha_f(\pi_{\mathcal{F}} I)$.*

(3) *The datum α_f is monotonic: If $F, F' \in \mathcal{F}$ and $F \subset F'$, then $\alpha_f(F) \geq \alpha_f(F')$.*

(4) *The collection $\mathcal{F}$ is σ-Carleson in that*

$$\sum_{F \in \mathcal{F} : F \subset S} \sigma(F) \leq 2\sigma(S), \qquad S \in \mathcal{D}. \tag{3.7}$$

(5) *We have the inequality*

$$\Big\| \sum_{F \in \mathcal{F}} \alpha_f(F) \cdot F \Big\|_\sigma \lesssim \|f\|_\sigma . \tag{3.8}$$

Proof
The first three properties are immediate from the construction. The fourth, the σ-Carleson property is seen this way. It suffices to check the property for $S \in \mathcal{F}$.

Now, the $\mathcal{F}$-children can be in $\mathcal{F}_{\text{energy}}(S)$, which satisfy

$$\sum_{F' \in \mathcal{F}_{\text{energy}}(S)} \sigma(F') \leq \frac{1}{10}\sigma(S).$$

Otherwise, note that by choice of $\alpha_f(\cdot)$, we have $\mathbb{E}_S^\sigma |f| \leq 2\alpha_f(S)$. These intervals F' satisfy $\mathbb{E}_{F'}^\sigma |f| \geq 10\alpha_f(S) \geq 5\mathbb{E}_S^\sigma |f|$. These intervals satisfy the display above with $1/10$ replaced by $1/5$. Hence, (3.7) holds.

For the final property, let $\mathcal{G} \subset \mathcal{F}$ be the subset at which the stopping values change. If $F \in \mathcal{F} - \mathcal{G}$, and G is the $\mathcal{G}$-parent of F, then $\alpha_f(F) = \alpha_f(G)$. Set

$$\Phi_G := \sum_{F \in \mathcal{F} : \pi_{\mathcal{G}} F = G} F.$$

Define $G_k := \{\Phi_G \geq 2^k\}$, for $k = 0, 1, \ldots$. The σ-Carleson property implies integrability of all orders in σ-measure of Φ_G. Using the third moment, we have $\sigma(G_k) \lesssim 2^{-3k}\sigma(G)$. Then, estimate

$$\begin{aligned}
\Big\| \sum_{F \in \mathcal{F}} \alpha_f(F) \cdot F \Big\|_\sigma^2 &= \Big\| \sum_{G \in \mathcal{G}} \alpha_f(G) \Phi_G \Big\|_\sigma^2 \\
&\leq \Big\| \sum_{k=0}^{\infty} (k+1)^{+1-1} \sum_{G \in \mathcal{G}} \alpha_f(G) 2^k \mathbf{1}_{G_k} \Big\|_\sigma^2 \\
&\stackrel{*}{\lesssim} \sum_{k=0}^{\infty} (k+1)^2 \Big\| \sum_{G \in \mathcal{G}} \alpha_f(G) 2^k \mathbf{1}_{G_k}(x) \Big\|_\sigma^2 \\
&\stackrel{**}{\lesssim} \sum_{k=0}^{\infty} (k+1)^2 \sum_{G \in \mathcal{G}} \alpha_f(G)^2 2^{2k} \sigma(G_k) \\
&\lesssim \sum_{G \in \mathcal{G}} \alpha_f(G)^2 \sigma(G) \lesssim \|Mf\|_\sigma^2 \lesssim \|f\|_\sigma^2.
\end{aligned}$$

Note that we have used Cauchy–Schwarz in k at the step marked by an $*$. In the step marked with $**$, for each point x, the nonzero summands are a (super)geometric sequence of scalars, so the square can be moved inside the sum. Finally, we use the estimate on the σ-measure of G_k, and compare to the maximal function Mf to complete the estimate. $\square$

We will use the notation

$$P_F^\sigma f := \sum_{I \in \mathcal{D} : \pi_{\mathcal{F}} I = F} \Delta_I^\sigma f, \quad F \in \mathcal{F},$$

and similarly for Q^w_F, but rather than use $\pi_{\mathcal{F}} J$ in the definition, we use $\dot{\pi}_{\mathcal{F}} J$, defined to be the minimal $F \in \mathcal{F}$ with $J \Subset F$. Without this alternate definition, some delicate case analysis would be forced upon us. The inequality (3.8) allows us to estimate

$$\begin{aligned}
&\sum_{F\in\mathcal{F}} \{\alpha_f(F)\sigma(F)^{1/2} + \|P^\sigma_F f\|_\sigma\}\|Q^w_F g\|_w \\
&\quad \le \Big[\sum_{F\in\mathcal{F}} \{\alpha_f(F)^2\sigma(F) + \|P^\sigma_F f\|^2_\sigma\} \times \sum_{F\in\mathcal{F}} \|Q^w_F g\|^2_w\Big]^{1/2} \\
&\quad \lesssim \|f\|_\sigma \|g\|_w. \qquad (3.9)
\end{aligned}$$

We will refer to this as the *quasiorthogonality* argument, and we remark that it only requires orthogonality of the projections $Q^w_F g$. It is very useful.

LEMMA 3.10
We have

$$\left|B^{\text{above}}(f,g) - B^{\text{above}}_{\mathcal{F}}(f,g)\right| \lesssim \mathscr{H}\|f\|_\sigma\|g\|_w,$$
$$\textit{where } B^{\text{above}}_{\mathcal{F}}(f,g) := \sum_{F\in\mathcal{F}} B^{\text{above}}(P^\sigma_F f, Q^w_F g).$$

Proof
We apply the functional energy inequality of Section 5. Observe that $f = \sum_{F\in\mathcal{F}} P^\sigma_F f$ and that

$$\sum_{J:J\Subset I_0} \Delta^w_J g = \sum_{F\in\mathcal{F}} Q^w_F g.$$

From the definition of $B^{\text{above}}(f,g)$, we can assume that g equals the sum above. Therefore,

$$B^{\text{above}}(f,g) = \sum_{F'\in\mathcal{F}} \sum_{F\in\mathcal{F}} B^{\text{above}}(P^\sigma_{F'} f, Q^w_F g).$$

In the sum above, we can also add the restriction that $F' \cap F \neq \emptyset$, for otherwise $B^{\text{above}}(P^\sigma_{F'} f, Q^w_F g) = 0$. For a pair of intervals $J \Subset I_J$, note that this implies that $J \Subset \pi_{\mathcal{F}} I$; that is, $\dot{\pi}_{\mathcal{F}} J \subset \pi_{\mathcal{F}} I$. Therefore, we can add the restriction $F \subset F'$. The case of $F' = F$ is the definition of $B^{\text{above}}_{\mathcal{F}}(f,g)$, so that it suffices to estimate

$$\sum_{\substack{F,F'\in\mathcal{F}\\ F'\supsetneq F}} B^{\text{above}}(P^\sigma_{F'} f, Q^w_F g). \qquad (3.11)$$

Observe that the functions $g_F := Q_F^w g$ are $\mathcal{F}$-adapted in the sense of Definition 5.1, and by construction $\mathcal{F}$ satisfies the Carleson measure condition (3.7). We take these steps to apply the functional energy inequality. The argument of the Hilbert transform is I_F, the child of I that contains F. Write $I_F = F + (I_F - F)$, and use linearity of H_σ. Note that by the standard martingale difference identity and the construction of stopping data,

$$\Big| \sum_{I : I \supsetneq F} \mathbb{E}^\sigma_{I_F} \Delta^\sigma_I f \Big| \lesssim \alpha_f(F), \quad F \in \mathcal{F}.$$

Hence, invoking interval testing,

$$\begin{aligned} \Big| \sum_{F \in \mathcal{F}} \sum_{I : I \supsetneq F} \mathbb{E}^\sigma_{I_F} \Delta^\sigma_I f \cdot \langle H_\sigma F, g_F \rangle_w \Big| &\lesssim \sum_{F \in \mathcal{F}} \alpha_f(F) \big| \langle H_\sigma F, g_F \rangle_w \big| \\ &\lesssim \mathcal{H} \sum_{F \in \mathcal{F}} \alpha_f(F) \sigma(F)^{1/2} \| g_F \|_w . \end{aligned}$$

Quasiorthogonality bounds this last expression.

For the second expression, when the argument of the Hilbert transform is $I_F - F$, first note that

$$\Big| \sum_{I : I \supsetneq F} \mathbb{E}^\sigma_{I_F} \Delta^\sigma_I f \cdot (I_F - F) \Big| \lesssim \Phi := \sum_{F' \in \mathcal{F}} \alpha_f(F') \cdot F', \quad F \in \mathcal{F}.$$

Therefore, by the definition of $\mathcal{F}$-adapted, the monotonicity property (4.3) applies, and yields

$$\Big| \sum_{I : I \supsetneq F} \mathbb{E}^\sigma_{I_F} \Delta^\sigma_I f \cdot \big\langle H_\sigma (I_F - F), g_F \big\rangle_w \Big| \lesssim \sum_{J \in \mathcal{J}^*(F)} P(\Phi\sigma, J) \Big\langle \frac{x}{|J|}, J \overline{g}_F \Big\rangle_w, \quad F \in \mathcal{F}.$$

Here, $\mathcal{J}^*(F)$ are the maximal good intervals $J \Subset F$, and $\overline{g}_F := \sum_{J \in \mathcal{J}(F) : J \Subset F} |\widehat{g}(J)| \cdot h^w_J$, so that every term has a positive inner product with x. The sum over $F \in \mathcal{F}$ of this last expression is controlled by functional energy, and by the property that $\|\Phi\|_\sigma \lesssim \|f\|_\sigma$. This completes the bound for (3.11). □

Proof of Theorem 3.5

By Lemma 3.10, it remains to control $B^{\text{above}}_{\mathcal{F}}(f, g)$. Keeping the quasiorthogonality argument in mind, we see that appropriate control on the individual summands is enough to control it. For each $F \in \mathcal{F}$, let $\mathcal{S}_F$ be the $\mathcal{F}$-children of F. Observe that the function

$$\big(C \alpha_f(F) \big)^{-1} P^\sigma_F f \tag{3.12}$$

is uniform on F with respect to $\mathcal{S}_F$, for an appropriate absolute constant C. Moreover, the function $Q^w_F g$ does not have any interval J in its Haar support strongly contained in an interval $S \in \mathcal{S}_F$. That is, it is weakly adapted to the function in (3.12). Therefore, by assumption,

$$\left| B^{\text{above}}(P^\sigma_F f, Q^w_F g) \right| \le \mathcal{L}\left\{\alpha_F(F)\sigma(F)^{1/2} + \|P^\sigma_F f\|_\sigma\right\} \|Q^w_F g\|_w.$$

The sum over $F \in \mathcal{F}$ of the right-hand side is bounded by the quasiorthogonality argument of (3.9). $\square$

4. Energy, monotonicity, and Poisson

Our theorem (Theorem 1.2) is particular to the Hilbert transform, and so depends upon special properties of it. They largely extend from the fact that the derivative of $-1/y$ is positive. The following monotonicity property for the Hilbert transform was observed in [7, Lemma 5.8], and is basic to the analysis of the functional energy inequality.

LEMMA 4.1 (Monotonicity property)
Let $K \supsetneq I$ be two intervals, and assume that σ does not have point masses at the end point of I. Then, for any function $g \in L^2(I, w)$, with w-integral zero, and $\beta > 2|K|$,

$$P\big(\sigma \cdot (K - I), I\big)\left\langle \frac{x}{|I|}, \overline{g}\right\rangle_w \lesssim \liminf_{\alpha \downarrow 0} \big\langle H_{\alpha,\beta}\big(\sigma(K-I)\big), \overline{g}\big\rangle_w. \tag{4.2}$$

Here, $\overline{g} = \sum_{J'} |\widehat{g}(J')| h^w_{J'}$ is a Haar multiplier applied to g. If J is a good interval, $J \Subset I$, then, for function $g \in L^2(J, w)$, with w-integral zero, and signed measures ν and μ supported on $K - I$, with $|\nu| \le \mu$, we have

$$\sup_{0<\alpha<\beta} \left| \langle H_{\alpha,\beta} \nu, g\rangle_w \right| \lesssim P(\mu, J)\left\langle \frac{x}{|J|}, \overline{g}\right\rangle_w. \tag{4.3}$$

The truncations enter into the formulation of the lemma, since they play a notable role here. We need this preparation.

LEMMA 4.4
Let I and J be two intervals which share an endpoint a, at which neither σ nor w have a point mass. Then,

$$\sup_{0<\alpha<\beta} \left| \langle H_{\alpha,\beta}\sigma I, J\rangle_w \right| \lesssim \mathcal{A}_2^{1/2} \sqrt{\sigma(I) w(J)}. \tag{4.5}$$

Proof
If $|I| \simeq |J|$, this inequality is the weak boundedness principle of [9, Section 2.2]. So, let us assume that $10|I| < |J|$. Then, it remains to bound

$$
\begin{aligned}
\big|\big\langle H_{\alpha,\beta}\sigma I, (J \setminus 10I)\big\rangle_w\big| &\le \sum_{n=11}^{\infty} \frac{\sigma(I) w(J \cap ((n+1)I \setminus nI))}{n|I|} \\
&\le \frac{\sigma(I)}{|I|^{1/2}} P(w,I)^{1/2} w(J)^{1/2} \lesssim \mathcal{A}_2^{1/2} \sqrt{\sigma(I) w(J)}.
\end{aligned}
$$

This depends upon obvious kernel bounds, and an application of Cauchy–Schwarz to derive the Poisson term above. □

Proof of Lemma 4.1

By linearity, it suffices to prove (4.2) in the case of $g = h_I^w$. The point is to separate the supports of the functions involved. Since I does not have a point mass at the end point of I, we have $\sigma(\lambda I \setminus I) \downarrow 0$ as $\lambda \downarrow 1$. It follows that we can fix a $\lambda > 1$ sufficiently small so that $P(\sigma(K-I), I) \simeq P(\sigma(K - \lambda I), I)$, and one more condition that we will come back to. Then, for $0 < \alpha < (1/2)(\lambda - 1)|I|$, we estimate as below, where x_I is the center of I,

$$
\begin{aligned}
&\big\langle H_{\alpha,\beta}\big(\sigma(K - \lambda I)\big), h_I^w\big\rangle_w \\
&\quad = \int_{K-\lambda I} \int_I \big\{K_{\alpha,\beta}(y - x) - K_{\alpha,\beta}(y - x_I)\big\} h_I^w(x) w(dx) \sigma(dy) \\
&\quad = \int_{K-\lambda I} \int_I \frac{x - x_I}{(y-x)(y-x_J)} h_I^w(x) w(dx) \sigma(dy) \\
&\quad \gtrsim P\big(\sigma(K - I), I\big) \Big\langle \frac{x - x_I}{|I|}, h_I^w \Big\rangle_w.
\end{aligned}
$$

We have subtracted the term, since h_I^w has integral zero, then we applied (2.3) with $C_{x,x_J,y} = 1$, as follows from our choices of α and β. Then, note that $(x - x_J) h_I^w \ge 0$, so that we can pull out the Poisson term. The last line follows by our selection of λ sufficiently close to 1. Then, the last condition needed is to select λ sufficiently close to 1 that, in view of (4.5),

$$
\sup_{\alpha,\beta} \big|\big\langle H_{\alpha,\beta}(\lambda I \setminus I), h_I^w\big\rangle_w\big| \lesssim \mathcal{A}_2^{1/2} \sqrt{\sigma(\lambda I \setminus I)} < c P\big(\sigma(K - I), I\big) \Big\langle \frac{x - x_I}{|I|}, h_I^w \Big\rangle_w.
$$

In the last line, $c > 0$ is an absolute constant. This completes the proof of (4.2).

Turn to (4.3). The estimate (2.3) applies:

$$
\begin{aligned}
\big|\langle H_{\alpha,\beta}\nu, g\rangle_w\big| &= \Big|\int_{K-I}\int_J \big\{K_{\alpha,\beta}(y - x) - K_{\alpha,\beta}(y - x_J)\big\} h_J^w(x) w(dx) \nu(dy)\Big| \\
&= \Big|\int_{K-I}\int_J C_{x,x_J,y} \frac{(x - x_J)}{(y - x)(y - x_J)} h_J^w(x) w(dx) \nu(dy)\Big|.
\end{aligned}
$$

But recall that $0 \le C_{x,x_J,y} \le 4$ and that it equals 1 for α sufficiently small. Moreover, $y-x$ and $y-x_J$ have the same sign, and $(x-x_J)h_J^w(x) \ge 0$. So an upper bound is obtained by passing from ν to μ:

$$\begin{aligned}\bigl|\langle H_{\alpha,\beta}\nu, g\rangle_w\bigr| &\le \int_{K-I}\int_J \frac{(x-x_J)}{(y-x)(y-x_J)} h_J^w(x)w(dx)\mu(dy)\\ &\simeq P(\mu,J)\Bigl\langle \frac{x}{|J|}, h_J^w\Bigr\rangle_w. \end{aligned}\qquad\square$$

The concept of *energy* is fundamental to the subject. For interval I, define

$$\begin{aligned}E(w,I)^2 &:= \mathbb{E}_I^{w(dx)}\mathbb{E}_I^{w(dx')}\frac{(x-x')^2}{|I|^2}\\ &= \frac{2}{w(I)}\sum_{J\subset I}\Bigl\langle \frac{x}{|I|}, h_J^w\Bigr\rangle_w^2. \end{aligned}\tag{4.6}$$

Now, consider the *energy constant*, the smallest constant $\mathcal{E}$ such that this condition holds, as presented or in its dual formulation. For all dyadic intervals I_0, all partitions $\mathcal{P}$ of I_0 into dyadic intervals, we have

$$\sum_{I\in\mathcal{P}} P(\sigma I_0, I)^2 E(w,I)^2 w(I) \le \mathcal{E}^2\sigma(I_0). \tag{4.7}$$

This was shown in [9, Proposition 2.11].

PROPOSITION 4.8
For a finite constant C_0,

$$\mathcal{E}^2 \le C_0\{\mathcal{A}_2^{1/2} + \mathcal{T}\}^2 = C_0\mathcal{H}^2.$$

We will always estimate $\mathcal{E}$ by $\mathcal{H}$. The proof is recalled here.

Proof
It suffices to consider the case of finite partitions $\mathcal{P}$ of I. We first prove a version of the energy inequality with *holes* in the argument of the Poisson. It follows from (4.2) that we can fix $0<\alpha<\beta$ such that

$$P\bigl(\sigma(I_0-I), I\bigr)^2 E(w,I)^2 w(I) \lesssim \bigl\| H_{\alpha,\beta}\bigl(\sigma(I_0-I)\bigr)\bigr\|_{L^2(I,\sigma)}^2, \quad I\in\mathcal{P}.$$

Then, using linearity and interval testing, we have

$$\sum_{I\in\mathcal{P}} \bigl\| H_{\alpha,\beta}(\sigma\cdot I_0)\bigr\|_{L^2(I,\sigma)}^2 \lesssim \bigl\| H_{\alpha,\beta}(\sigma\cdot I_0)\bigr\|_{L^2(I,\sigma)}^2 \lesssim \mathcal{H}^2\sigma(I_0)$$

and

$$\sum_{I\in\mathcal{P}} \left\| H_{\alpha,\beta}(\sigma\cdot I)\right\|^2_{L^2(I,\sigma)} \lesssim \mathcal{H}^2 \sum_{I\in\mathcal{P}} \sigma(I) \lesssim \mathcal{H}^2 \sigma(I_0).$$

Then, by the A_2 bound, we have $P(\sigma\cdot I, I)^2 E(w,I)^2 w(I) \lesssim \sigma(I)$, which we can sum over the partition. This completes the proof. □

One should keep in mind that the concept of energy is related to the tails of the Hilbert transform. The energy inequality, and its multiscale extension to the functional energy inequality, show that the control of the tails is very subtle in this problem.

We also need the following elementary Poisson estimate from [29]; used occasionally in this argument, it is crucial to the proof of Lemma 3.1.

LEMMA 4.9
Suppose that $J \Subset I \subset I_0$, and that J is good. Then

$$|J|^{2\epsilon-1} P\big(\sigma(I_0-I), J\big) \lesssim |I|^{2\epsilon-1} P\big(\sigma(I_0-I), I\big). \tag{4.10}$$

Proof
We have $\operatorname{dist}(J, I_0-I) \ge |J|^{\epsilon}|I|^{1-\epsilon}$, so that for any $x\in I_0-I$, we have

$$\frac{|J|^{2\epsilon}}{(|J|+\operatorname{dist}(x,J))^2} \lesssim \frac{|I|^{2\epsilon}}{(|I|+\operatorname{dist}(x,I))^2}.$$

Integrating this last expression, it follows that

$$\begin{aligned} |J|^{2\epsilon-1} P\big(\sigma\cdot(I_0-I), J\big) &= |J|^{2\epsilon-1}\int_{I_0-I} \frac{|J|}{(|J|+\operatorname{dist}(x,J))^2}\,d\sigma \\ &\lesssim |I|^{2\epsilon}\int_{I_0-I}\frac{1}{(|J|+\operatorname{dist}(x,J))^2}\,d\sigma. \end{aligned}$$

And this proves the inequality. □

5. The functional energy inequality

We state an important multiscale extension of the energy inequality (4.7).

Definition 5.1
Let $\mathcal{F}$ be a collection of dyadic intervals. A collection of (good) functions $\{g_F\}_{F\in\mathcal{F}}$ in $L^2(w)$ is said to be *$\mathcal{F}$-adapted* if, for all $F\in\mathcal{F}$, the Haar support of the function g_F is contained in $\{J : \dot{\pi}_{\mathcal{F}} J = F\}$.

Definition 5.2
Let $\mathscr{F}$ be the smallest constant in the inequality below, or its dual form. The inequality holds for all nonnegative $h \in L^2(\sigma)$, all σ-Carleson collections $\mathcal{F}$, and all $\mathcal{F}$-adapted collections $\{g_F\}_{F\in\mathcal{F}}$:

$$\sum_{F\in\mathcal{F}} \sum_{J^*\in\mathcal{J}^*(F)} P(h\sigma, J^*) \Big| \Big\langle \frac{x}{|J^*|}, g_F J^* \Big\rangle_w \Big| \le \mathscr{F} \|h\|_\sigma \Big[\sum_{F\in\mathcal{F}} \|g_F\|_w^2 \Big]^{1/2}.$$

Here $\mathcal{J}^*(F)$ consists of the *maximal* good intervals $J \Subset F$. Note that the estimate is universal in h and $\mathcal{F}$, separately.

This constant was identified in [7], and is herein shown to be necessary from the A_2 and interval testing inequalities. Recall the definition of $\mathscr{H}$ in (1.5).

THEOREM 5.3
Assume that $\mathscr{F}$ satisfies (3.7); then, $\mathscr{F} \lesssim \mathscr{H}$.

The first step in the proof is the domination of the constant $\mathscr{F}$ by the best constant in a certain two-weight inequality for the Poisson operator, with the weights being determined by w and σ in a particular way. This is the decisive step, since there is a two-weight inequality for the Poisson operator proved by one of us in [27, Theorem 2]. It reduces the full norm inequality to simpler testing conditions, which are in turn controlled by the A_2 and Hilbert transform testing conditions.

5.1. The two-weight Poisson inequality
Consider the weight

$$\mu \equiv \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F)} \Big\| P^w_{F,J} \frac{x}{|J|} \Big\|_w^2 \cdot \delta_{(x_J, |J|)}.$$

Here, $P^w_{F,J} := \sum_{J' : J' \subset J, \dot{\pi}_{\mathcal{F}} J = F} \Delta^w_{J'}$. We can replace x by $x - c$ for any choice of c we wish; the projection is unchanged. And δ_q denotes a Dirac unit mass at a point q in the upper half-plane $\mathbb{R}^2_+$. We prove the two-weight inequality for the Poisson integral

$$\big\| \mathbb{P}(h\sigma) \big\|_{L^2(\mathbb{R}^2_+, \mu)} \lesssim \mathscr{H} \|h\|_\sigma,$$

for all nonnegative h. Above, $\mathbb{P}(\cdot)$ denotes the Poisson extension operator to the upper half-plane, so that in particular

$$\big\| \mathbb{P}(h\sigma) \big\|^2_{L^2(\mathbb{R}^2_+, \mu)} = \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F)} \mathbb{P}(h\sigma)\big(x_J, |J|\big)^2 \Big\| P^w_{F,J} \frac{x}{|J|} \Big\|_w^2,$$

where x_J is the center of the interval J. The proof of Theorem 5.3 follows by duality.

Phrasing things in this way brings a significant advantage. The characterization of the two-weight inequality for the Poisson operator (see [27]) reduces the full norm inequality above to these testing inequalities. For any dyadic interval $I \in \mathcal{D}$ we have

$$\int_{\mathbb{R}^2_+} \mathbb{P}(\sigma \cdot I)^2 \, d\mu(x,t) \lesssim \mathcal{H}^2 \sigma(I), \tag{5.4}$$

$$\int_{\mathbb{R}} \mathbb{P}^*(t\widehat{I}\mu)^2 \sigma(dx) \lesssim \mathcal{A}_2 \int_{\widehat{I}} t^2 \, d\mu(x,t), \tag{5.5}$$

where $\widehat{I} = I \times [0,|I|]$ is the box over I in the upper half-plane and where $\mathbb{P}^*$ is the dual Poisson operator

$$\mathbb{P}^*(t\widehat{I}\mu) = \int_{\widehat{I}} \frac{t^2}{t^2 + |x-y|^2} \mu(dy,dt).$$

One should keep in mind that the intervals I are restricted to be in our fixed dyadic grid, a reduction allowed as the integrations on the left-hand side in (5.4) and (5.5) are done over the entire space, either $\mathbb{R}^2_+$ or $\mathbb{R}$. (Goodness of the intervals I above is not needed.) This reduction is critical to the analysis below.

Remark 5.6
A gap in the proof of the Poisson inequality in [27, p. 542] can be fixed as in [28] or [8].

5.2. *The Poisson testing inequality: The core*

This section is concerned with a part of inequality (5.4). Restrict the integral on the left-hand side to the set $\widehat{I} \subset \mathbb{R}^2_+$:

$$\int_{\widehat{I}} \mathbb{P}(\sigma \cdot I)^2 \, d\mu(x,t) \lesssim \mathcal{H}^2 \sigma(I).$$

Since $(x_J, |J|) \in \widehat{I}$ if and only if $J \subset I$, we have

$$\int_{\widehat{I}} \mathbb{P}(\sigma \cdot I)(x,t)^2 \, d\mu(x,t) = \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F): J \subset I} \mathbb{P}(\sigma \cdot I)\big(x_J, |J|\big)^2 \left\| P^w_{F,J} \frac{x}{|J|} \right\|_w^2 .$$

For each J,

$$\left\| P^w_{F,J} \frac{x}{|J|} \right\|_w^2 \leq \int_J \left| \frac{x - \mathbb{E}^w_J x}{|J|} \right|^2 dw(x) = 2E(w,J)^2 w(J) \leq 2w(J). \tag{5.7}$$

Let $\mathcal{F}_0$ be the maximal $F \in \mathcal{F}$ which are strictly contained in I, and let $\mathcal{J}^\sharp$ be those dyadic J such that $(x_J, |J|)$ is in the support of μ, but has no parent in $\mathcal{F}_0$.

These intervals are necessarily disjoint. Observe that by (5.7) and the energy inequality,

$$\begin{aligned}\sum_{J\in\mathcal{J}^\sharp}\mathbb{P}(\sigma F)\big(x_J,|J|\big)^2\mu\big(x_J,|J|\big)&\lesssim\sum_{J\in\mathcal{J}^\sharp}P(\sigma\cdot F,J)^2E(w,J)^2w(J)\\&\lesssim\mathcal{H}^2\sigma(F).\end{aligned}\tag{5.8}$$

We claim that

$$\sum_{F\in\mathcal{F}_0}\int_{\hat F}\mathbb{P}\big(\sigma(I\setminus F)\big)(x,t)^2\,d\mu(x,t)\lesssim\mathcal{H}\sigma(I).\tag{5.9}$$

This is sufficient, since

$$\begin{aligned}&\int_{\widehat I}\mathbb{P}(\sigma\cdot I)(x,t)^2\,d\mu(x,t)\\&\quad\lesssim\ \text{LHS(5.8)}\ +\ \text{LHS(5.9)}\ +\sum_{F\in\mathcal{F}_0}\int_{\hat F}\mathbb{P}(\sigma\cdot F)(x,t)^2\,d\mu(x,t)\\&\quad\lesssim\mathcal{H}^2\sigma(I)+\sum_{F\in\mathcal{F}_0}\int_{\hat F}\mathbb{P}(\sigma\cdot F)(x,t)^2\,d\mu(x,t).\end{aligned}$$

The individual terms in the last sum are set up for a recursive application of this inequality. Due to the Carleson condition (3.7), this recursion will finish the proof.

It remains to prove (5.9), which is another instance of the energy inequality. For an interval $F_0\in\mathcal{F}_0$, and $F\in\mathcal{F}$ strictly contained in F_0, each interval $J\in\mathcal{J}^*(F)$ is contained in some $J_0\in\mathcal{J}^*(F_0)$. Then, the intervals $F\in\mathcal{F}$ are not good, but J and J_0 are good, hence

$$\begin{aligned}\mathbb{P}\big(\sigma(I\setminus F_0)\big)\big(x_J,|J|\big)^2\mu\big(x_J,|J|\big)&=\Big[\int_{I\setminus F_0}\frac{|J|}{|J|^2+|x-x_J|^2}\Big]^2\Big\|P^w_{F,J}\frac{x}{|J|}\Big\|_w^2\\&=\Big[\int_{I\setminus F_0}\frac{1}{|J|^2+|x-x_J|^2}\Big]^2\|P^w_{F,J}x\|_w^2\\&\lesssim\Big[\int_{I\setminus F_0}\frac{|J_0|}{|J_0|^2+|x-x_{J_0}|^2}\Big]^2\Big\|P^w_{F,J}\frac{x}{|J_0|}\Big\|_w^2.\end{aligned}$$

This follows from goodness. For $x\in I\setminus F_0$,

$$|J|^2+|x-x_J|^2\geq|x-x_J|^2\geq|x-x_{J_0}|^2\geq|J_0|^\epsilon|F_0|^{1-\epsilon}.$$

But then, we can add the projections $P^w_{F,J}$, due to orthogonality, and use (5.7) again to see that

$$\begin{aligned}
&\sum_{\substack{F\in\mathcal{F}\\ F\subset F_0}}\sum_{\substack{J\in\mathcal{J}^*(F)\\ J\subset J_0}} \mathbb{P}\big(\sigma(I\setminus F_0)\big)\big(x_J,|J|\big)^2\mu\big(x_J,|J|\big)\\
&\qquad\lesssim \mathbb{P}(\sigma\cdot I)\big(x_{J_0},|J_0|\big)^2\sum_{\substack{F\in\mathcal{F}\\ F\subset F_0}}\sum_{\substack{J\in\mathcal{J}^*(F)\\ J\subset J_0}}\Big\|P^w_{F,J}\frac{x}{|J_0|}\Big\|_w^2\\
&\qquad\lesssim \mathbb{P}(\sigma\cdot I)\big(x_{J_0},|J_0|\big)^2E(w,J_0)^2w(J_0).
\end{aligned}$$

The sum over $F_0\in\mathcal{F}_0$, and $J_0\in\mathcal{J}^*(F_0)$ is controlled by the energy inequality. This completes the proof of (5.9).

5.3. *The Poisson testing inequality: The remainder*

Now we turn to proving the following estimate for the global part of the first testing condition (5.4):

$$\int_{\mathbb{R}^2_+-\widehat{I}}\mathbb{P}(\sigma\cdot I)^2\,d\mu\lesssim\mathcal{A}_2\sigma(I).$$

Decompose the integral on the left-hand side into four terms. With F_J the unique $F\in\mathcal{F}$ with $J\in\mathcal{J}^*(F)$, and using (5.7),

$$\begin{aligned}
&\int_{\mathbb{R}^2_+-\widehat{I}}\mathbb{P}(\sigma\cdot I)^2\,d\mu\\
&\quad=\sum_{J:(x_J,|J|)\in\mathbb{R}^2_+-\widehat{I}}\mathbb{P}(\sigma\cdot I)\big(x_J,|J|\big)^2\Big\|P^w_{F_J,J}\frac{x}{|J|}\Big\|_w^2\\
&\quad\le\Big\{\sum_{\substack{J:J\cap 3I=\emptyset\\ |J|\le|I|}}+\sum_{J:J\subset 3I-I}+\sum_{\substack{J:J\cap I=\emptyset\\ |J|>|I|}}+\sum_{J:J\supsetneqq I}\Big\}\mathbb{P}(\sigma\cdot I)\big(x_J,|J|\big)^2w(J)\\
&\quad=A+B+C+D.
\end{aligned}$$

Decompose term A according to the length of J and its distance from I, to obtain

$$\begin{aligned}
A&\lesssim\sum_{n=0}^{\infty}\sum_{k=1}^{\infty}\sum_{\substack{J:J\subset 3^{k+1}I-3^kI\\ |J|=2^{-n}|I|}}\Big(\frac{2^{-n}|I|}{\operatorname{dist}(J,I)^2}\sigma(I)\Big)^2w(J)\\
&\lesssim\sum_{n=0}^{\infty}2^{-2n}\sum_{k=1}^{\infty}\frac{|I|^2\sigma(I)w(3^{k+1}I-3^kI)}{|3^kI|^4}\sigma(I)\\
&\lesssim\sum_{n=0}^{\infty}2^{-2n}\sum_{k=1}^{\infty}3^{-2k}\Big\{\frac{\sigma(3^{k+1}I)w(3^{k+1}I)}{|3^kI|^2}\Big\}\sigma(I)\lesssim\mathcal{A}_2\sigma(I).
\end{aligned}$$

Decompose term B according to the length of J and then use the Poisson inequality (4.10), available to use because of goodness of intervals J. We then obtain

$$\begin{aligned} B &\lesssim \sum_{n=0}^{\infty} \sum_{\substack{J : J \subset 3I - I \\ |J| = 2^{-n}}} 2^{-n(2-4\epsilon)} \frac{\sigma(I)^2}{|I|^2} w(J) \\ &\lesssim \sum_{n=0}^{\infty} 2^{-n(2-4\epsilon)} \frac{\sigma(3I) w(3I)}{|3I|^2} \sigma(I) \lesssim \mathscr{A}_2 \sigma(I). \end{aligned}$$

For term C, for $n = 1, 2, \dots$, set $\mathscr{J}_n$ to be those good dyadic intervals J with $|J| > |I|$, $J \cap I = \emptyset$ and

$$(n-1)|J| \leq \operatorname{dist}(I, J) < n|J|.$$

These intervals have bounded overlaps. Indeed, suppose that $J_1 \subsetneq \cdots \subsetneq J_r$ are all members for $\mathscr{J}_1$. Then, by goodness,

$$\begin{aligned} \operatorname{dist}(J_1, I) \geq \operatorname{dist}(J_r, I) &\geq (n-1) 2^r |J_1| + \operatorname{dist}(J_1, \partial J_r) \\ &\geq \left\{ (n-1) 2^r + 2^{r(1-\epsilon)} \right\} |J_1|, \end{aligned}$$

which is a contradiction to membership in $\mathscr{J}_n$. Restricting the sum to intervals in $\mathscr{J}_n$, we have

$$\begin{aligned} \sum_{J \in \mathscr{J}_n} \mathbb{P}(\sigma \cdot I) \big(x_J, |J| \big)^2 w(J) &\lesssim \sigma(I)^2 \sum_{J \in \mathscr{J}_n} \frac{w(J)}{n^4 |J|^2} \\ &\lesssim \frac{\sigma(I)^2}{|I|} \sum_{J \in \mathscr{J}_n} \frac{w(J) \cdot |I|}{n^4 |J|^2} \\ &\lesssim \frac{\sigma(I)}{n^2} \cdot \frac{\sigma(I)}{|I|} P(w, I) \lesssim \mathscr{A}_2 \frac{\sigma(I)}{n^2}. \end{aligned}$$

And this is summable in $n \in \mathbb{N}$.

In the last term D, all the intervals J contain I. Note that

$$\begin{aligned} \sum_{J : J \supsetneq I} \mathbb{P}(\sigma \cdot I) \big(x_J, |J| \big)^2 w(J) &\lesssim \sigma(I)^2 \sum_{J : J \supsetneq I} \frac{w(J)}{|J|^2} \\ &\lesssim \sigma(I) \cdot \frac{\sigma(I)}{|I|} \sum_{J : J \supsetneq I} \frac{w(J) \cdot |I|}{|J|^2} \\ &\lesssim \sigma(I) \cdot \frac{\sigma(I)}{|I|} P(w, I) \lesssim \mathscr{A}_2 \sigma(I). \end{aligned}$$

5.4. The dual Poisson testing inequality

We are considering (5.5). Note that there is a power of t on both sides and that the expressions on the two sides of this inequality are

$$\int_{\widehat{I}} t^2 \mu(dx,dt) = \sum_{F\in\mathcal{F}} \sum_{\substack{J\in\mathcal{J}^*(F)\\ J\subset I}} \|P^w_{F,J}x\|_w^2,$$

$$\mathbb{P}^*(t\widehat{I}\mu)(x) = \sum_{F\in\mathcal{F}} \sum_{\substack{J\in\mathcal{J}^*(F)\\ J\subset I}} \frac{\|P^w_{F,J}x\|_w^2}{|J|^2+|x-x_J|^2}.$$

We are to dominate $\|\mathbb{P}^*(t\widehat{I}\mu)\|_\sigma^2$ by the first expression above. The squared norm will be the sum over integers s of T_s below, in which the relative lengths of J and J' are fixed by s. Suppressing the requirement that $J, J'\subset I$,

$$\begin{aligned}T_s &:= \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F)} \sum_{F'\in\mathcal{F}} \sum_{\substack{J'\in\mathcal{J}^*(F)\\ |J'|=2^{-s}|J|}} \int \frac{\|P^w_{F,J}x\|_w^2}{|J|^2+|x-x_J|^2}\cdot \frac{\|P^w_{F',J'}x\|_w^2}{|J'|^2+|x-x_{J'}|^2}\,d\sigma\\ &\le M_s \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F)} \|P^w_{F,J}x\|_w^2,\end{aligned}$$

where

$$M_s \equiv \sup_{F\in\mathcal{F}} \sup_{J\in\mathcal{J}^*(F)} \sum_{F'\in\mathcal{F}} \sum_{\substack{J'\in\mathcal{J}^*(F)\\ |J'|=2^{-s}|J|}} \int \frac{1}{|J|^2+|x-x_J|^2}\cdot \frac{w(J')\cdot|J'|^2}{|J'|^2+|x-x_{J'}|^2}\,d\sigma.$$

The estimate (5.7) has been used in the definition of M_s. We claim the term M_s is at most a constant times $\mathcal{A}_2 2^{-s}$, and it is here that the full Poisson A_2 condition is used.

Fix J, and let $n\in\mathbb{N}$ be the integer chosen so that $(n-1)|J|\le \operatorname{dist}(J,J')\le n|J|$. Estimate the integral in the definition of M_s by

$$\frac{w(J')}{|J'|}\int \frac{|J'|^2}{|J|^2+|x-x_J|^2}\cdot\frac{|J'|}{|J'|^2+|x-x_{J'}|^2}\,d\sigma \lesssim \mathcal{A}_2 2^{-2s}.$$

This estimate is adequate for $n=0,1,2$. Then estimate the sum over J' as

$$\sum_{F'\in\mathcal{F}} \sum_{\substack{J'\in\mathcal{J}^*(F'):|J'|=2^{-s}|J|\\ (n-1)|J|\le\operatorname{dist}(J,J')\le n|J|}} 2^{-2s}\lesssim 2^{-s},$$

because the relative lengths of J and J' are fixed, and each J' is in at most one $\mathcal{J}^*(F)$.

For the case of $n \geq 3$, restrict J' to be to the right of J, and let $t_n = x_J + x_{J'}/2$, so that $|x_J - t_n|, |x_{J'} - t_n| \simeq n|J|$. First, estimate the integral in the definition of M_s on the interval $[t_n, \infty)$:

$$\frac{w(J')}{|J'|} \int_{t_n}^{\infty} \frac{|J'|^2}{|J|^2 + |x - x_J|^2} \cdot \frac{|J'|}{|J'|^2 + |x - x_{J'}|^2} \, d\sigma \lesssim \mathscr{A}_2 \frac{2^{-2s}}{n^2}.$$

Then estimate the sum over J' as follows:

$$\sum_{F' \in \mathcal{F}} \sum_{\substack{J' \in \mathcal{J}^{\epsilon}(F') : |J'| = 2^{-s}|J| \\ (n-1)|J| \leq \operatorname{dist}(J, J') \leq n|J|}} \frac{2^{-2s}}{n^2} \lesssim \frac{2^{-s}}{n^2}.$$

This is clearly summable in $n \geq 4$.

Now, estimate on the integral on the interval $(-\infty, t_n)$:

$$\begin{aligned} &\frac{w(J')}{|J'|} \int_{-\infty}^{t_n} \frac{|J'|^2}{|J|^2 + |x - x_J|^2} \cdot \frac{|J'|}{|J'|^2 + |x - x_{J'}|^2} \, d\sigma \\ &\quad = 2^{-2s} \frac{w(J')}{|J|} \int_{-\infty}^{t_n} \frac{|J|}{|J|^2 + |x - x_J|^2} \cdot \frac{|J|^2}{|J'|^2 + |x - x_{J'}|^2} \, d\sigma \\ &\quad \lesssim 2^{-2s} \frac{w(J')}{n^2 |J|} P(\sigma, J). \end{aligned}$$

Drop the term with the geometric decay in s, and sum over n and J' to see that

$$\sum_{n=4}^{\infty} \sum_{F' \in \mathcal{F}} \sum_{\substack{J' \in \mathcal{J}^{\epsilon}(F') : |J'| = 2^{-s}|J| \\ (n-1)|J| \leq \operatorname{dist}(J, J') \leq n|J|}} \frac{w(J')}{n^2 |J|} P(\sigma, I) \lesssim P(w, J) P(\sigma, J) \lesssim \mathscr{A}_2.$$

Here, we have appealed to the full Poisson A_2 condition. This completes the control of the dual Poisson testing condition.

Acknowledgments. The authors benefited from a stimulating conference on two-weight inequalities in October 2011 at the American Institute of Mathematics in Palo Alto, California, for which they express their appreciation. The reviewing process led to many improvements in this paper, including a streamlining of the main contribution, for which we thank the referees.

References

[1] G. DAVID and J.-L. JOURNÉ, *A boundedness criterion for generalized Calderón-Zygmund operators*, Ann. of Math. (2) **120** (1984), 371–397. MR 0763911. DOI 10.2307/2006946. *(2797)*

[2] V. P. HAVIN and N. K. NIKOLSKI, eds., *Linear and Complex Analysis: Problem Book 3, Part I*, Lecture Notes in Math. **1573**, Springer, Berlin, 1994, 318–319. MR 1334345. *(2797)*

[3] T. P. HYTÖNEN, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. (2) **175** (2012), 1473–1506. MR 2912709. DOI 10.4007/annals.2012.175.3.9. *(2797)*

[4] T. P. HYTÖNEN, M. T. LACEY, H. MARTIKAINEN, T. ORPONEN, M. C. REGUERA, E. T. SAWYER, and I. URIARTE-TUERO, *Weak and strong type estimates for maximal truncations of Calderón-Zygmund operators on A_p weighted spaces*, J. Anal. Math. **118** (2012), 177–220. MR 2993026. DOI 10.1007/s11854-012-0033-3. *(2797)*

[5] M. T. LACEY, *Two weight inequality for the Hilbert transform: A real variable characterization, II*, Duke Math. J. **163** (2014), 2821–2840. DOI 10.1215/00127094-2826799. *(2796)*

[6] M. T. LACEY, S. PETERMICHL, and M. C. REGUERA, *Sharp A_2 inequality for Haar shift operators*, Math. Ann. **348** (2010), 127–141. MR 2657437. DOI 10.1007/s00208-009-0473-y. *(2797)*

[7] M. T. LACEY, E. T. SAWYER, C.-Y. SHEN, and I. URIARTE-TUERO, *The two weight inequality for Hilbert transform, coronas, and energy conditions*, preprint, arXiv:1108.2319v1 [math.CA]. *(2797, 2802, 2803, 2807, 2811)*

[8] M. T. LACEY, E. T. SAWYER, and I. URIARTE-TUERO, *A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure*, Anal. PDE **5** (2012), 1–60. MR 2957550. DOI 10.2140/apde.2012.5.1. *(2812)*

[9] ———, *A two weight inequality for the Hilbert transform assuming an energy hypothesis*, J. Funct. Anal. **263** (2012), 305–363. MR 2923415. DOI 10.1016/j.jfa.2012.04.019. *(2796, 2797, 2800, 2807, 2809)*

[10] Y. I. LYUBARSKII and K. SEIP, *Complete interpolating sequences for Paley-Wiener spaces and Muckenhoupt's (A_p) condition*, Rev. Mat. Iberoam. **13** (1997), 361–376. MR 1617649. DOI 10.4171/RMI/224. *(2797)*

[11] B. MUCKENHOUPT, *Hardy's inequality with weights*, Studia Math. **44** (1972), 31–38. MR 0311856. *(2797)*

[12] ———, *Weighted norm inequalities for the Hardy maximal function*, Trans. Amer. Math. Soc. **165** (1972), 207–226. MR 0293384. *(2797)*

[13] F. NAZAROV, *A counterexample to Sarason's conjecture*, preprint, http://www.math.msu.edu/~fedja/prepr.html (accessed 12 September 2014). *(2797)*

[14] F. NAZAROV, S. TREIL, and A. VOLBERG, *Cauchy integral and Calderón-Zygmund operators on nonhomogeneous spaces*, Int. Math. Res. Not. IMRN **1997**, no. 15, 703–726. MR 1470373. DOI 10.1155/S1073792897000469. *(2797, 2800)*

[15] ———, *Accretive system Tb-theorems on nonhomogeneous spaces*, Duke Math. J. **113** (2002), 259–312. MR 1909219. DOI 10.1215/S0012-7094-02-11323-4. *(2797, 2800)*

[16] ———, *The Tb-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), 151–239. MR 1998349. DOI 10.1007/BF02392690. *(2797, 2800)*

[17] ———, *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Res. Lett. **15** (2008), 583–597. MR 2407233. DOI 10.4310/MRL.2008.v15.n3.a16. *(2797)*

[18] ———, *Two weight estimate for the Hilbert transform and corona decomposition for non-doubling measures*, preprint, arXiv:1003.1596v1 [math.AP]. *(2797, 2800, 2802)*

[19] F. NAZAROV and A. VOLBERG, *The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces K_θ*, J. Anal. Math. **87** (2002), 385–414. MR 1945290. DOI 10.1007/BF02868482. *(2797)*

[20] N. NIKOLSKI and S. TREIL, *Linear resolvent growth of rank one perturbation of a unitary operator does not imply its similarity to a normal operator*, J. Anal. Math. **87** (2002), 415–431. MR 1945291. DOI 10.1007/BF02868483. *(2797)*

[21] C. PÉREZ, S. TREIL, and A. VOLBERG, *On A_2 conjecture and corona decomposition of weights*, preprint, arXiv:1006.2630v1 [math.AP]. *(2797)*

[22] ———, *Sharp weighted estimates for dyadic shifts and the A_2 conjecture*, preprint, arXiv:1010.0755v2 [math.CA]. *(2797)*

[23] A. POLTORATSKI and D. SARASON, “Aleksandrov-Clark measures” in *Recent Advances in Operator-related Function Theory (Dublin, 2004)*, Contemp. Math. **393**, Amer. Math. Soc., Providence, 2006, 1–14. MR 2198367. DOI 10.1090/conm/393/07366. *(2797)*

[24] M. C. REGUERA and J. SCURRY, *On joint estimates for maximal functions and singular integrals on weighted spaces*, Proc. Amer. Math. Soc. **141** (2013), 1705–1717. MR 3020857. DOI 10.1090/S0002-9939-2012-11474-1. *(2796)*

[25] D. SARASON, “Exposed points in H^1, II” in *Topics in Operator Theory: Ernst D. Hellinger Memorial Volume*, Oper. Theory Adv. Appl. **48**, Birkhäuser, Basel, 1990, 333–347. MR 1207406. *(2797)*

[26] E. T. SAWYER, *A characterization of a two-weight norm inequality for maximal operators*, Studia Math. **75** (1982), 1–11. MR 0676801. *(2797)*

[27] ———, *A characterization of two weight norm inequalities for fractional and Poisson integrals*, Trans. Amer. Math. Soc. **308**, no. 2 (1988), 533–545. MR 0930072. DOI 10.2307/2001090. *(2797, 2811, 2812)*

[28] E. T. SAWYER and R. L. WHEEDEN, *Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces*, Amer. J. Math. **114** (1992), 813–874. MR 1175693. DOI 10.2307/2374799. *(2812)*

[29] A. VOLBERG, *Calderón-Zygmund Capacities and Operators on Nonhomogeneous Spaces*, CBMS Reg. Conf. Ser. Math. **100**, Amer. Math. Soc., Providence, 2003. MR 2019058. *(2796, 2797, 2800, 2802, 2810)*

Lacey

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA;

lacey@math.gatech.edu

Sawyer
Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1; sawyer@mcmaster.ca

Shen
Department of Mathematics, National Central University, Chungli, 32054, Taiwan; chunyshen@gmail.com

Uriarte-Tuero
Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, USA; ignacio@math.msu.edu