THE TWO WEIGHT INEQUALITY FOR THE HILBERT TRANSFORM: A PRIMER

MICHAEL T LACEY

Abstract. Given a pair of weights \( w, \sigma \), the two weight inequality for the Hilbert transform is of the form
\[
\|H(\alpha f)\|_{L^2(w)} \lesssim \|f\|_{L^2(\sigma)}.
\]
Recent work of Lacey-Sawyer-Shen-Uriarte-Tuero and Lacey have established a conjecture of Nazarov-Treil-Volberg, giving a real-variable characterization of which pairs of weights this inequality holds, provided the pair of weights do not share a common point mass. In this paper, the characterization is proved, collecting details from across several papers; compactness is characterized; all relevant estimates are proved; counterexamples are details; and areas of application are indicated.

1. Introduction

By a weight we mean a non-negative Borel locally finite measure, typically on \( \mathbb{R} \). We consider the two weight inequality for the Hilbert transform for a pair of weights \( w, \sigma \) on \( \mathbb{R} \):
\[
\sup_{0 < \alpha < \beta < \infty} \|H_{\alpha, \beta}(f \cdot \sigma)\|_{L^2(w)} \leq N\|f\|_{L^2(\sigma)}.
\]

## Contents

1. Introduction 1
2. Preliminaries 9
3. Necessary Conditions 12
4. Global to Local Reduction 21
5. The Stopping Form 29
6. Elementary Estimates 40
7. Functional Energy Inequality 45
8. Two Weight Poisson Inequality 49
9. Compact Operators 58
10. Proof under the Pivotal Assumption 62
11. Example Weights 68
12. Applications of the Main Inequality 75
References 79
Here, \( N \) denotes the best constant in the inequality. And, \( H_{\alpha,\beta} \nu(x) \), for signed measure \( \nu \), denotes a standard truncation of the usual Hilbert transform, given by

\[
(1.2) \quad H_{\alpha,\beta} \nu(x) := \int_{\alpha < |x-y| < \beta} \frac{\nu(dy)}{y-x}.
\]

The inequality is phrased this way as the familiar principal value need not exist in the settings we are interested in. Below, however, we will suppress the truncation parameters \( \alpha \) and \( \beta \), understanding that the relevant inequalities are uniform over \( 0 < \alpha < \beta < \infty \). Thus the norm inequality above will be written as \( \|H(\sigma f)\|_w \leq N \|f\|_\sigma \). (There are a few points where the truncations will reappear, in the course of the argument.) In areas of application, there is no \textit{a priori} conditions placed upon the weights that can arise in this question, forcing us to consider the question in this form.

The central question is then a real-variable characterization of the inequality (1.1). In the special case that the pair of weights \( \sigma \) and \( \omega \) do not share a common point mass, this was supplied in two papers, one of Lacey-Sawyer-Shen-Uriarte-Tuero [22], and another of the present author [16], answering a beautiful conjecture of Nazarov-Treil-Volberg [58].

\textbf{Theorem 1.3.} Suppose that for all \( x \in \mathbb{R} \), \( \sigma(\{x\}) \cdot \omega(\{x\}) = 0 \) for the pair of weights \( \sigma, \omega \). Define two positive constants \( A_2 \) and \( T \) as the best constants in the inequalities below, uniform over intervals \( I \).

\[
(1.4) \quad P(\sigma, I) \cdot P(\omega, I) \leq A_2,
\]

\[
(1.5) \quad \int_I H(\sigma 1_I)^2 \, d\omega \leq T^2 \sigma(I),
\]

\[
\int_I H(\omega 1_I)^2 \, d\sigma \leq T^2 \omega(I).
\]

There holds \( N \simeq \mathcal{H} := A_2^{1/2} + T \).

The first condition is an extension of the Muckenhoupt \( A_2 \) condition to a Poisson condition. The exact Poisson extension of \( \sigma \) to the upper half-plane is not needed, rather we use the approximation below, which is roughly the Poisson extension evaluated at the center of \( I \), and up into the half-plane the length of \( I \), see Figure 1.

\[
P(\sigma, I) := \int_{\mathbb{R}} \frac{|I|}{(|I| + \text{dist}(x,I))^2} \sigma(dx).
\]

The remaining conditions are referred to as the Sawyer-type testing conditions, as he first introduced these conditions into the two weight setting in his fundamental papers on the maximal function [49], and later the fractional and Poisson integral operators [50]. It is well-known that the \( A_2 \) condition is necessary for the two weight inequality, and it is obvious that the testing conditions are necessary. Thus, the substance of the result concerns the sufficiency of the \( A_2 \) and testing inequalities for the norm inequality.

This Theorem is a central result in the non-homogeneous harmonic analysis, as founded in a sequence of influential papers of Nazarov-Treil-Volberg [32–34].

The proof of the theorem is very involved, encompassing arguments and points of view that were spread across several papers [16, 21, 22, 35]. An important additional result, Sawyer’s two weight inequality for the Poisson integral [50], is not well-represented in the literature. Finally, the interest in the two weight inequality is well-motivated by applications to operator theory, model spaces, and spectral theory, themselves spread across additional papers.
The value of $P(\sigma, I)$ is approximately the Poisson extension of $\sigma$ evaluated at point in the upper half-plane given by the center of $I$, and the length of $I$.

The point of this paper is to
(a) state and prove the Theorem, in all detail, including a complete proof of Sawyer’s two weight Poisson inequality;
(b) supply the characterization of compactness of $H_\sigma$, which is a bit more complicated than one might expect;
(c) give the proof under the influential pivotal condition, which serves to highlight where the difficulties arise in the general case;
(d) collect relevant, explicit, counterexamples;
(e) give complements and extensions of the theorem, and the proof techniques;
(f) and point to areas of applications.

Sections proceed directly towards proofs, but many conclude with some context and discussion. The proof is entirely elementary, assuming only the well known facts about martingale differences.

Throughout, we will assume that the pair of weights do not share a common point mass. This assumption allows a wide range of ‘mollifications’ of the Hilbert transform, as is elegantly shown in [28]. While the question, as we have formulated it, makes sense in the setting in common point masses, this assumption introduces a number of subtle, even mysterious, issues, and so we leave it aside.

1.1. Compactness. A characterization of compactness follows, stated for weights without point masses. ‘Compactness’ in this context requires elaboration, since we understand the norm boundedness of the Hilbert transform to mean a uniform bound on the family of truncations. Again, a uniformity is understood. One definition of compactness of an operator $A$ between two Hilbert spaces, $A : \mathcal{H}_1 \mapsto \mathcal{H}_2$, is that for each $\varepsilon > 0$ there is a finite dimensional operator $B : \mathcal{H}_1 \mapsto \mathcal{H}_2$ such that $\|A - B\| \leq \varepsilon$. Make this definition uniform:

**Definition 1.6.** By $H_\sigma : L^2(\sigma) \mapsto L^2(w)$ is compact it is understood that for all $\varepsilon > 0$, there are $N_\varepsilon \in \mathbb{N}$, and $0 < \alpha_\varepsilon < \beta_\varepsilon < \infty$ so that for all choices of truncations $0 < \alpha < \beta$, there is a $B_{\alpha,\beta} : L^2(\sigma) \mapsto L^w$, with norm bounded by a fixed constant times $N$, and the dimension of the range of $B_{\alpha,\beta}$ is at most $N_\varepsilon$, and, finally that

$$\sup_{0 < \alpha < \beta} \|H_{\alpha,\beta}(\sigma f) - B_{\alpha,\beta}(\sigma f)\|_w \leq \varepsilon \|f\|_\sigma.$$

Moreover, if $0 < \alpha < \beta < \alpha_\varepsilon$ or $\beta_\varepsilon < \alpha < \beta$, then $B_{\alpha,\beta} \equiv 0$. 

Theorem 1.7. Let \( \sigma, w \) be two weights on \( \mathbb{R} \) with neither having point masses, and the the two weight inequality (1.1) holds. Then, \( H_\sigma : L^2(\sigma) \mapsto L^2(\sigma) \) is compact if and only if these conditions hold: For \( \Lambda > 0 \), write \( \sigma = \sigma_\Lambda^0 + \sigma_\Lambda^1 \), where \( \sigma_\Lambda^0 = \sigma 1_{[-\Lambda,\Lambda]} \), and similarly for \( w \).

\[
\lim_{\Lambda \uparrow \infty} \sup_{1 \in \mathbb{R} \setminus [-\Lambda,\Lambda]} P(\sigma_\Lambda^0, I) \cdot P(w_\Lambda^0, I) = 0, \quad \Lambda > 0
\]

\[
\lim_{\Lambda \uparrow \infty} \sup_{1 \in \mathbb{R} \setminus [-\Lambda,\Lambda]} P(\sigma_\Lambda^1, I) \cdot P(w_\Lambda^1, I) = 0
\]

\[
\lim_{\Lambda \uparrow \infty} \sup_{1 \in \mathbb{R} \setminus [-\Lambda,\Lambda]} \sigma(I)^{-1} \int_I |H_\sigma 1_I|^2 \, dw = 0, \quad \Lambda > 0,
\]

\[
\lim_{\Lambda \uparrow \infty} \sup_{1 \in \mathbb{R} \setminus [-\Lambda,\Lambda]} \sigma(I)^{-1} \int_I |H_\sigma 1_I|^2 \, dw = 0
\]

There are two more conditions dual to (1.11) and (1.12), And there are three more conditions dual to (1.10) which must hold: Interchange the roles of the \( \sigma \) and \( w \) in the half-Poisson condition as written, and then form the corresponding limits as \( \Lambda \to -\infty \) to get the last two conditions.

The characterization is in terms of ‘vanishing’ conditions: (a) over all small scales, uniformly as \( \Lambda \uparrow \infty \); (b) over all scales ‘at infinity’; and (c) uniformly over certain symmetric half-Poisson products. If the weights are compactly supported on \([-1,1]\), say, then the conditions reduce to (1.8) and (1.11) with its dual, and \( \Lambda = 1 \).

We have assumed that the weights do not have point masses, which simplifies the statement of the theorem. Two examples in \( \S 9.4 \) show that the ‘vanishing’ conditions above are no longer true in the case of point masses, and therefore complicating the characterization of compactness in this case.

1.2. An Overview of the Proof. The result is specific to the Hilbert transform, meaning that particular properties of this transform must guide the proof. The elementary examples of these are the monotonicity principle, Lemma 3.13, valid for all pairs of weights, and then the energy inequality, Lemma 3.17, valid under the assumption of interval testing and the \( \Lambda_2 \) condition. There is a third critical property, the functional energy inequality. These properties are a last vestige of positivity: The kernel \( \frac{1}{|y|} \) is monotone increasing on \( \mathbb{R} \setminus \{0\} \). This feature will deliver to us the energy inequality; finding it, and unlocking its secrets is the key to the proof.

The proof strategy is outlined in Figure 2. One begins with the bilinear form \( \langle H_\sigma f, g \rangle_w \). The passage to the ‘triangular forms’ in Lemma 4.3 is a rather standard step in many T1-type theorems. The Calderón-Zygmund stopping data defined in \( \S 4 \) is a the foundational tool. It (a) controls the values of certain telescoping sums of martingale differences; (b) regularizes the weights, from the point of view of the energy inequality; and (c) allows the the use of the quasi-orthogonality argument, an important simplification. The triangular forms are of a ‘local’ and a ‘global’ form, and have dual forms as well. There are two steps in the analysis, a ‘global to local’ reduction in \( \S 4 \), and an analysis of the ‘stopping form’ in the \( \S 5 \). These steps are familiar to experts in the T1 theorem, but carrying them out, however, necessarily depends upon novel techniques.

The stopping data is essential ‘global to local reduction’ in Theorem 4.11. A simple appeal to the testing condition, allows an application of the monotonicity principle to rephrase the inequality in this Theorem as a certain two-weight inequality for the Poisson integral. In this inequality, the Poisson
integral maps functions on $\mathbb{R}$ to those on $\mathbb{R}^2_+$. The weight on $\mathbb{R}$ is, say, $\sigma$. The weight on $\mathbb{R}^2_+$ is then derived from $w$ in a specific fashion. The resulting inequality, called the functional energy inequality, is a deep extension of the energy inequality. It then very fortunate that Sawyer has proved the two weight inequality for the Poisson integral §8, and only testing conditions need be verified. These are then reduced to the $A_2$ conditions, and the energy inequality itself. Notably, it is only here that the full Poisson $A_2$ condition is used.

The local term is then dominated by the analysis of the stopping form (5.1). This is again a familiar object, to experts in $T_1$ theorem, addressed by ad hoc off-diagonal estimates, which absolutely do not apply in the current context. Control of the irregularities of the weights is now the main point, complicated by the fact that the stopping form is not intrinsically defined. A notion of ‘size’ is introduced—it serves as an approximate of the operator norm of the stopping form, and again is most naturally defined in terms of a measure on $\mathbb{R}^2_+$, derived from one of the two given weights. The size Lemma, Lemma 5.6, makes the approximation property: Using the structure of the derived weight, one can decompose a stopping form into constituent parts. Those of large size have a simpler form, which allows one to estimate their operator norm by size. What is left has smaller size, and so one can recurse.

Some readers will have noticed that a very common set of objects, Carleson measures, are not mentioned, and indeed, they do not appear in the proof at all.\footnote{The functional energy inequality can be seen as a certain weighted Carleson measure: It shows that the Poisson operator embeds $L^2(\mathbb{R},\sigma)$ into $L^2(\mathbb{R}^2_+,\mu)$, where $\mu$ is derived from $w$ and $\sigma$ in a particular way.} The wide spread prevalence of Carleson measures in $T_1$ theorems can be traced to two facts, first that associated paraproduct operators are the principle obstacle to a simple proof, and second, the paraproduct operators have an essentially canonical form. In this theorem, neither of these facts hold, and so we have abandoned the notions of Carleson measures and paraproducts.

Carleson measures are also used to, indirectly, control the sums of martingale differences. To achieve this, we will use stopping data, as described in §4. It reappears at different points of the argument, for more or less the same reason.
Figure 3. For $0 < \epsilon < 1$, the function $w(x) = |x|^{1-\epsilon}$ is an $A_2$ weight. It and the dual weight $\sigma(x) = |x|^\epsilon - 1$ are graphed above. One can check that $[w]_{A_2} \approx \epsilon^{-1}$.

1.3. The $A_2$ Theory. The classical case of an $A_2$ weight corresponds to the case of $w(dx) = w(x)dx$, and $w(x) > 0$ a.e. Moreover, the weight $\sigma$ also has density given by $\sigma(x) := w(x)^{-1}$. It is assumed that both $w$ and $\sigma$ are locally integrable, so that they are both weights. See Figure 3. Note that $w(x) \cdot \sigma(x) \equiv 1$. The Muckenhoupt $A_2$ condition asserts that this same equality approximately holds, uniformly over location and scale.

$$[w]_{A_2} := \sup_I \frac{w(I)}{|I|} \cdot \frac{\sigma(I)}{|I|} < \infty.$$ 

These are ‘simple’ averages. This condition is equivalent to the uniform norm bound on $L^2(w)$ for the class of simple averaging operators

$$f \mapsto \frac{1}{|I|} \int_I f \, dx \cdot 1_I, \quad I \text{ is an interval.}$$

From this condition flows a rich theory, including the boundedness of all Calderón-Zygmund operators. The classical result of Hunt-Muckenhoupt-Wheeden [11] states that $w$ is in $A_2$ if and only if the Hilbert transform maps $L^2(w)$ to $L^2(w)$. By a basic change of variables argument, first noted by Sawyer [49], this is equivalent to $H_\sigma$ mapping $L^2(\sigma)$ to $L^2(w)$. Stefanie Petermichl [41] quantified the Hunt-Muckenhoupt-Wheeden theorem as follows.

**Theorem A.** A weight $w \in A_2$ if and only if $H$ is bounded from $L^2(w)$ to $L^2(w)$, and moreover $N \sim [w]_{A_2}$.

To place this result in the context of our main result, it is classical and easy to see that the Poisson $A_2$ characteristic satisfies $A_2 \leq [w]_{A_2}^2$. And, using the remarkable Haar shift representation of the Hilbert transform due to Petermichl [40], one can check that the testing condition satisfies $T \leq [w]_{A_2}$. This is what Petermichl’s original proof did. A more conceptual approach to this estimate was given in [17]. All existing proofs of Petermichl’s Theorem (see [12, 26]) depend ultimately on known Lebesgue measure estimates for the Hilbert transform, or closely related operators. For instance, [17] uses the weak-$L^1(dx)$ bound for Haar shift operators. Estimates of these type are irrelevant for the two weight theorem.

It is perhaps worth emphasizing that the powerful Haar shift technique of Petermichl, even with its impressive extension by Hytönen [12], seems to be of little use in the general two weight problem. There are two obstacles: Firstly, in order to use it, one must essentially have control on a Haar shift operator, independently of how the grid defining the shift is defined. The resulting condition on the pair of weights
is more subtle than the two weight inequality for the Hilbert transform. Secondly, one should recover the energy inequality of Lemma 3.17. But, the energy of any fixed Haar shift is zero, and indeed, the two weight inequality for Haar shift operators [36] has just a few difficulties in its proof.

By the $A_2$ Theorem, it is meant the linear in $A_2$ bound for all Calderón-Zygmund operators. This result, pursued by many, and established by Hytönen [12], has many points of contact with the subject of this note. But, we refer the reader to [14] and references there in for more information.

In the $A_2$ theory, it is essential that $w(x) > 0$ a.e. Suppose one relaxes this condition to $w(x)$ is positive on a measurable set $E \subset \mathbb{R}$, and define $\sigma(x)$ to be supported on $E$, and equal to $w(x)^{-1}$. One can then ask if the Hilbert transform is bounded for this pair of weights, and Theorem 1.3 applies here.

This question is an instance of the non-homogeneous $A_2$ theory advocated by A. Volberg. One can hope that specificity in the way the weights are prescribed could introduce some additional simplifications in the characterization of the two weight inequality in this setting. But, none has yet been found.

1.4. The Individual Two Weight Problem. Given an operator $T$, the individual $L^p$ two weight inequality for $T$ is the inequality

$$(1.13) \quad \|Tf\|_{L^p(w)} \leq N_T \|f\|_{L^p(\sigma)}.$$ 

Here and throughout we use the notation $Tf := T(\sigma f)$. We understand that $T$ applied to a signed measure $\sigma \cdot f$ should make sense. And, the inequality above is the preferred form of the inequality as duality is expressed in the natural way: The inequality (1.13) is equivalent to

$$(\|T^* g\|_{L^{p'}(\sigma)} \leq N_T \|g\|_{L^{p'}(w)}).$$ 

The question is then to characterize the pairs of weights for which (1.13) holds.

This specificity of the question is of interest for a few canonical operators, ones for which the corresponding two weight inequality will naturally present itself. The leading examples of this are, for positive operators, the Hardy operator by Muckenhoupt [29], the maximal function, Sawyer’s Theorem of 1981 [49] and Sawyer’s 1988 theorem for the fractional integrals [50]. It is noteworthy that the two weight inequalities for the Hardy and the Poisson integral are used in the proof of our main theorem.

It is interesting to that this is not only a chronological list, but it also reflects the depth of the results as well. The Hardy operator is easiest, characterized by an ‘$A_2$-type condition,’ as recalled in Theorem F. It was Sawyer’s insight, however, that the maximal function characterization requires a testing condition. The fractional integrals are harder still. For the sake of comparison, let us state a special case of the result for the fractional integrals in one dimension.\textsuperscript{2} One can also compare to Theorem G for the Poisson integral. Both results give a characterization in terms of testing conditions. And, while we state just one case of the general result, one should note that there is no Sobolev condition imposed on the $L^p$ indices.

**Theorem B.** For two weights $w, \sigma$, and $0 < \alpha < 1$, the operator $R_{\sigma}f(x) := \int f(x-y) \frac{\sigma(dy)}{|y|^{\alpha}}$ maps $L^2(\sigma)$ to $L^2(w)$ if and only if the testing inequalities below hold:

$$\int I R_{\sigma}(1_I)^2 \, dw \leq \mathcal{T}^2 \sigma(I), \quad \int I R_w(1_I)^2 \, d\sigma \leq \mathcal{T}^2 w(I).$$

Moreover the norm of the operator is equivalent to $\mathcal{T}$, the best constant in the inequalities above.

\textsuperscript{2}Besides Sawyer’s results, one should also consult Cascante-Ortega-Verbitsky [6].
The analysis of the individual two weight inequality for positive operators is much simpler, as is the case of dyadic operators. For certain non-positive dyadic operators, see the result of Nazarov-Treil-Volberg [36]. All of these results have found significant interest, due to the Haar shift operators of Petermichl [40], the remarkable median inequality of Lerner [25], and the Hytönen representation theorem [12].

The Hilbert transform is the first non-positive continuous operator for which the individual two weight problem has been solved. And, one would only ever expect that the solution would be of interest (or even possible) for a few canonical choices of operators, such as Hilbert, Cauchy and Riesz transforms. Foundational to the solution for the Hilbert transform is the monotonicity of the kernel. No other canonical choice will satisfy such a simple condition.

The individual two weight question makes sense for any $1 < p < \infty$, and there are characterizations in this, and other off-diagonal cases for positive operators. But, for singular integrals and even their dyadic analogs, it does not seem that there will be useful characterizations in the case of $p \neq 2$. See [19] for some results in this case, for maximal truncations of singular integrals.

1.5. The Hilbert Transform. The two weight inequality for the Hilbert transform was addressed as early as 1976 by Muckenhoupt and Wheeden [30]. But, it received much wider recognition as an important problem with the 1988 work of Sarason [47]. The latter was part of important sequence of investigations that identified de Branges spaces as an essential tool in operator theory. His question concerning the composition of Toeplitz operators, see §12.1, was raised therein, and advertised again in [48]. This question related the individual two weight problem for the Hilbert transform to a profound question from operator theory.

While not stated in the language of the Hilbert transform, Sarason wrote that it was ‘tempting’ to conjecture that the full Poisson $A_2$ condition would be sufficient for the two weight inequality. In an important development, F. Nazarov [31] showed that this was not the case. The two weight problem was seen to be important to Model spaces, namely certain embedding questions for Model spaces can be realized as a two weight inequality for the Hilbert transform. In particular, a more delicate counterexample was developed by Nazarov-Volberg [37] to disprove a conjectured characterization of the Carleson measures for a model space. The Nazarov counterexample was also used by Nikol’skiǐ-Treil [39], in the context of spectral theory.

The Nazarov counterexample is by way of a Bellman function approach. In §11, we give an explicit example. It is worth noting that in Sarason’s question, the weights have a density $|f|^2$, for analytic $f$, and the subharmonicity could be an important part of the problem. But, in the context of model spaces, completely arbitrary measures can arise. In §11, one of the weights is uniform measure on a Cantor set.

Nazarov-Treil-Volberg were creating the field of non-homogeneous Harmonic Analysis, in a series of ground-breaking papers [32–34]. Their work, and a revitalization of the perspective of Eric Sawyer from the 1980’s, lead them to conjecture the characterization proved in this paper. Moreover, their influential proof strategy, devised in [35,58], lead to a verification of the conjecture in the case that both weights were doubling. This paper uses their strategy, with several additional features. At the same time, their approach is generic, in that it applies to general Calderón-Zygmund operators. Specific properties of the Hilbert transform had to be used in the characterization. These properties were identified in [16,20–22], and the more precise description of what was accomplished at each stage is spread out throughout the paper.

---

3In particular, they noted that the simple $A_2$ condition was not sufficient for the boundedness of the Hilbert transform, and conjectured that half-Poisson $A_2$ conditions would be sufficient, an indication of the powerful sway held by the Muckenhoupt $A_2$ condition in the early years of the weighted theory.
and negative values of $h$. This is the familiar martingale difference equality, and so we will refer to $h$ as a martingale difference.

\[ E \]

It implies the familiar telescoping identity

\[
\{ \sigma[I_{0}] \} = \{ \sigma[I_{0}] \} \cup \{ \sigma[I_{0}] \}.
\]

2. Preliminaries

2.1. Dyadic Grids and Haar Functions. A grid is a collection $D$ of closed intervals so that for all $I, J \in D$, $I \cap J = \emptyset$, $I, J$. Further say that $D$ is a dyadic grid if for all integers $n$, the collection $\{ I \in D : |I| = 2^n \}$ partitions $\mathbb{R}$, aside from the endpoints of the intervals.

For a sub collection $F$ of a dyadic grid $D$, set $\pi_{F} I$ to be the minimal element of $F$ that contains $I$; $I$ need not be a member of $F$. Set $\pi_{F} I$ to be the minimal member of $F$ that strictly contains $I$, inductively define $\pi_{F}^{-1} I = \pi_{F}(\pi_{F}^{-1} I)$.

Say that the collection $D$ is admissible for weight $\sigma$ if $\sigma$ does not have a point mass at any endpoint of an interval $I \in D$.

2.2. Haar Functions. Let $D$ be admissible for $\sigma$ be a weight on $\mathbb{R}$. If $I \in D$ is such that $\sigma$ assigns non-zero weight to both children of $I$, the associated Haar function is chosen to have a non-negative inner product with the independent variable, $\langle x, h_I(x) \rangle_{\sigma} \geq 0$, a convenient choice due to the central role of the energy inequality, (3.18).

\[
h_I(x) := \sqrt{\frac{\sigma(I_{-}) \sigma(I_{+})}{\sigma(I)}} \left( \frac{I_{+}(x)}{\sigma(I_{+})} - \frac{I_{-}(x)}{\sigma(I_{-})} \right).
\]

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 $\sigma$-integral zero. If $\sigma$ is supported only on one child of $I$, then we set $h_I \equiv 0$.

For any dyadic interval $I_0$ with $\sigma(I_0) > 0$, the non-zero functions among $\{ \sigma(I_0)^{-1/2} I_0 \} \cup \{ h_I^\sigma : I \in D, I \subset I_0 \}$ form an orthonormal basis for $L^2(I_0, \sigma)$. We will use the notation $L_0^\sigma(I_0, \sigma)$ for the subspace of $L^2(I_0, \sigma)$ of functions with mean zero. It has orthonormal basis consisting of the non-zero functions in $\{ h_I^\sigma : I \in D, I \subset I_0 \}$. These are familiar properties. But, another familiar property, that the positive and negative values of $h_I^\sigma$ are comparable in absolute value, fails in a dramatic fashion for non-doubling measures. See Figure 4.

We will use the notations $\hat{f}(I) = \langle f, h_I^\sigma \rangle$, as well as the equality below, holding for those $I$ with $h_I^\sigma \neq 0$.

\[
\Delta_I^\sigma f = \langle f, h_I^\sigma \rangle h_I^\sigma = I_+ E_I^\sigma f + I_- E_I^\sigma f - I E_I^\sigma f.
\]

This 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 $E_I^\sigma f = \sum_{1 \leq |J|} E_J^\sigma \Delta_J^\sigma f$.

The Haar support of a function $f \in L^2(\sigma)$ is the collection $\{ I : \hat{f}(I) \neq 0 \}$. 

\[\text{Figure 4. Two Haar functions. For the left function, the weight is nearly equally distributed between the two halves of the interval, in sharp contrast to the function on the right, in which the weight on the right half is much larger than on the left.}\]
2.3. **Random Dyadic Grids.** Let \( \hat{D} \) be the standard dyadic grid in \( \mathbb{R} \), thus all intervals \([0, 2^n]\) for \( n \in \mathbb{N} \) are in \( \hat{D} \). A random dyadic grid \( D \) is specified by \( \omega = \{ \omega_n \} \in \{0, 1\}^\mathbb{Z} \), and the elements are

\[
I = \hat{I} + \omega := \hat{I} + \sum_{n: 2^{-n} < |I|} 2^{-n} \omega_n, \quad \hat{I} \in \hat{D}.
\]

The natural uniform probability measure \( \mathbb{P} \) is placed upon \( \{0, 1\}^\mathbb{Z} \).

Fix \( 0 < \varepsilon < 1 \) and \( r \in \mathbb{N} \). An interval \( I \in D \) is said to be \((\varepsilon, r)\)-good if for all intervals \( J \in D \) with \(|J| \geq 2^{r-1}|I|\), the distance from \( \partial J \) and either child of \( I \) is at least \(|I|^\varepsilon |J|^{1-\varepsilon} \). Otherwise \( I \) is said to be \((\varepsilon, r)\)-bad. These are the basic properties of this definition.

**Proposition 2.2.** These three properties hold.

1. The property of \( I = \hat{I} + \omega \) being \((\varepsilon, r)\)-good only depends upon \( \omega \) and \(|I|\).
2. \( P_{\text{good}} := \mathbb{P}(I \text{ is } (\varepsilon, r)\text{-good}) \) is independent of \( I \).
3. \( P_{\text{bad}} := 1 - P_{\text{good}} \leq \varepsilon^{-1} 2^{-\varepsilon r} \).

**Proof.** An interval \( I = \hat{I} + \omega \) is equally likely to be the left or right half of its parent \( \pi_D I \), depending only on \( \omega_n \), where \(|I| = 2^n\). Similarly, \( I \) is equally likely to be any one of the \( 2^t \) potential positions in \( \pi_D I \), and its exact position is determined by \( \{\omega_n, \ldots, \omega_{n+t-1}\} \). This proves the first two claims.

For the last, if \( I \) is bad, then for some \( t > r \), there holds \( \text{dist}(I, \partial \pi_D I) \leq 2^{(1-\varepsilon)t}|I| \). For this to happen, it is necessary that the numbers \( \{\omega_s : n + [(1-\varepsilon)t] < u \leq n + t - 1\} \) all be equal, and hence are either all \( 0 \) or all \( 1 \). This clearly proves that

\[
P_{\text{bad}} \leq \sum_{t=r+1}^{\infty} 2^{1-(t-[(1-\varepsilon)t])} \leq \varepsilon^{-1} 2^{-\varepsilon r}.
\]

This elementary proposition is used in the following fundamental way. Fix two weights \( w, \sigma \). With probability one, a random \( D \) is admissible for both \( w \) and \( \sigma \). Indeed, the collection of points that are point masses for one of the two weights is a fixed countable collection of points. And any fixed point has probability zero of being an endpoint of an interval in \( D \). Hence, we can, with probability one, define the Haar basis adapted to these two weights. Write the identity operator on \( L^2(\sigma) \) by

\[
P_{\text{good}}^\sigma f + P_{\text{bad}}^\sigma f \quad \text{where} \quad P_{\text{good}}^\sigma f := \sum_{I \in D: I \text{ is } (\varepsilon, r)\text{-good}} (f, h_I^\sigma)_{\sigma} h_I^\sigma.
\]

Use the same notation for the weight \( w \).

**Proposition 2.3.** There holds

\[
\mathbb{E}\|P_{\text{bad}}^\sigma f\|_{\sigma}^2 \leq \varepsilon^{-1} 2^{-\varepsilon r} \|f\|_{\sigma}^2.
\]

**Proof.** The location of \( I \) and the property of \( I \) being bad are independent, hence

\[
\mathbb{E}\|P_{\text{bad}}^\sigma f\|_{\sigma}^2 = \mathbb{E} \sum_{I \in D} 1_{I \text{ is bad}} \hat{f}(I)^2 = P_{\text{bad}} \mathbb{E} \sum_{I \in D} \hat{f}(I)^2 = \|f\|_{\sigma}^2
\]

and then the proposition follows. \(\square\)
Lemma 2.4. For any \( 0 < \varepsilon < 1 \), there is a choice of \( r \in \mathbb{N} \) sufficiently large so that this holds. Let \( w, \sigma \) be a pair of weights for which the constant \( \mathcal{H} \) is finite, and suppose there holds uniformly over admissible dyadic grids \( D \),

\[
\langle H_\sigma p_\text{good}^\sigma f, p_\text{good}^w g \rangle_w \leq \mathcal{H} \|f\|_\sigma \|g\|_w,
\]

then, the best constant (1.1) satisfies \( N \leq \mathcal{H} \).

Proof. Recall that the two weight norm inequality is uniform over all truncations \( 0 < \tau < 1 \) as in (1.2). Restricting \( \tau_0 < \tau < 1 \), for some fixed positive \( \tau_0 \), the \( A_2 \) condition then implies a bound \( N_{\tau_0} \) on the operators \( H_\tau \). Indeed, for \( x_0 \in \mathbb{R} \), let \( I_0 = (x_0 - \tau_0/2, x_0 + \tau_0/2) \). Then, for constants \( C_0 \) that only depend upon \( \tau_0 \), for all \( \tau_0 < \tau < 1 \),

\[
\int_{I_0} \left| \int_{\tau < |x-y| < \tau^{-1}} f(y) \frac{\sigma(dy)}{y-x} \right|^2 w(dx)
\leq C_0 \int_{\tau_0/2 < |x_0-y| < 2\tau_0^{-1}} f(y)^2 \sigma(dy) \int_{\mathbb{R} - I_0} p_{\tau_0}(y)^2 \sigma(dy) \frac{w(I_0)}{|I_0|}
\leq C_0 A_2 \int_{\tau_0/2 < |x_0-y| < 2\tau_0^{-1}} f(y)^2 \sigma(dy).
\]

There is no need to have an effective control on the constant \( C_0 \). This is summed over \( x_0 \in (\tau_0/2, \tau_0) \), to complete the proof the finiteness of \( N_{\tau_0} \), assuming the \( A_2 \) condition.

The argument below shows that \( N_{\tau_0} \leq \mathcal{H} \). Use Proposition 2.3 on the good and bad projections, as written and the same version for \( L^2(w) \).

\[
|\langle H_\sigma f, g \rangle_w| \leq \mathbb{E} \{|\langle H_\sigma p_\text{good}^\sigma f, p_\text{good}^w g \rangle_w| + |\langle H_\sigma p_\text{good}^\sigma f, p_\text{bad}^w g \rangle_w| + |\langle H_\sigma p_\text{bad}^\sigma f, p_\text{good}^w g \rangle_w| + |\langle H_\sigma p_\text{bad}^\sigma f, p_\text{bad}^w g \rangle_w|\}.
\]

The first term is controlled by the assumption (2.5), and the remaining terms are controlled by the finiteness of \( N_{\tau_0} \) and average-norm estimate on the bad projection. By appropriate selection of \( f \in L^2(\sigma) \) and \( g \in L^2(w) \), there holds

\[
N_{\tau_0} \leq \mathcal{H} + \varepsilon^{-1} 2^{-\varepsilon r/2} N_{\tau_0}.
\]

For any fixed \( \varepsilon \), we can take \( r \geq \varepsilon^{-1} \log \varepsilon^{-1} \), so that the second term can be absorbed into the left hand side. \( \square \)

2.4. Context and Discussion.

2.4.1. The random grid method was pioneered in [33], and is a critical tool in non-homogeneous analysis [58], where the weights need not be doubling. It has a broader set of uses, as witnessed by a powerful representation of a general Calderón-Zygmund operator as a rapidly convergent sum of dyadic operators due to Hytönen [12].

2.4.2. The parameterization of the grids used here follows Hytönen [13], but the statistics of this parameterization are those of the random shift in Nazarov-Treil-Volberg [32, 33].
3. Necessary Conditions

Herein, we take up the necessity of the $A_2$ condition from the norm inequality. Following that is the monotonocity property, an essential property of the Hilbert transform, and then showing the necessity of the energy inequality from the $A_2$ and interval testing condition. The energy inequality is foundational to the proof, and is elaborated in the section on functional energy §7.

3.1. The $A_2$ Condition. The $A_2$ condition has different forms, and so we clarify the language associated with the $A_2$ condition here. The simple $A_2$ condition is

$$\sup_I \frac{\sigma(I)}{|I|} \cdot \frac{w(I)}{|I|},$$

the supremum formed over all intervals $I$. This reduces to the classical Muckenhoupt condition if $w(dx) = w(x) dx$, where $w(x) > 0$ a.e., and $\sigma(dx) = w(x)^{-1} dx$. Next, are the half-Poisson conditions:

$$\sup_I P(\sigma, I) \frac{w(I)}{|I|}.$$

For much of the proof of sufficiency, we only need the half-Poisson condition, as written above, and in its dual form. Finally there is the (full) Poisson $A_2$ condition of (1.4), which we will use in the proof of the functional energy inequality.

We verify that the Poisson $A_2$ condition (1.4) is necessary for the two weight inequality (1.1).

**Proposition 3.1.** Assume that the pair of weights do not share a common point mass, and that the norm inequality (1.1) holds. Then, the $A_2$ condition (1.4) holds.

**Proof.** Assume that $w$ and $\sigma$ are supported on a compact interval $[-\beta, \beta]$, so that various integrals below are necessarily finite. In addition, the $\beta$ parameter will be incorporated into truncations of the Hilbert transform, using the definition (1.2). It is convenient to use the notation

$$(3.2) \quad p_1(x)^2 := \frac{|I|}{(|I| + \text{dist}(x, I))^2},$$

so that $P(\sigma, I) = \|p_1\|^2_{L^2(\sigma)}$.

A general inequality is derived. For $y < x$, and any interval $I$, there holds

$$|I|(x - y) = |I|((x - x_I) - (y - x_I)) \leq (|I| + |x - x_I|) \cdot (|I| + |y - x_I|)$$

Thus,

$$p_1(x) p_1(y) \leq \frac{1}{x - y}, \quad y < x.$$ 

We use this inequality, incorporating an arbitrary $a \in \mathbb{R}$, and a choice of $0 < \alpha < \beta$, to see that

$$p_1(x) \int_{(-\infty,a)} p_1(y)^2 \sigma(dy) \leq \int_{(-\infty,a)} \frac{1}{x - y} p_1(y) \sigma(dy)$$

$$= -H_{\alpha, \beta} (\sigma p_1 1_{(-\infty,a)})(x), \quad \alpha + 2\alpha < x.$$ 

Here, we are using the truncations as given in (1.2), and taking $\beta' = |a| + \beta$, where recall that the weights are supported on $[-\beta, \beta]$. It is important that $\alpha + 2\alpha < x$. We emphasize that this argument only depends upon the canonical value argument of (3.7), which holds for many choices of truncations.
From the assumed norm inequality, uniform over all truncations, there holds
\[ \int_{[a+a,\infty)} p_1(x)^2 \left[ \int_{(-\infty,a)} p_1(y)^2 \sigma(dy) \right]^2 \, dw \leq \int_{[a+a,\infty)} H_{\alpha,\beta}'(\sigma p_11_{(-\infty,a)})^2 \, dw \]
(3.3)
\[ \leq N^2 P(\sigma 1_{(-\infty,a)}, I). \]

And, rearranging, there holds
\[ P(\sigma 1_{(-\infty,a)}, I) \cdot P(w 1_{[a+a,\infty)}, I) \leq N^2. \]
This holds for all \( \alpha > 0 \), whence taking \( \alpha \downarrow 0 \),
(3.4) \[ P(\sigma 1_{(-\infty,a)}, I) \cdot P(w 1_{(a,\infty)}, I) \leq N^2. \]
This inequality holds for all intervals \( I \) and \( \alpha \), noting that the two weights are restricted to complementary half-lines. Clearly, it holds with the roles of \( w \) and \( \sigma \) reversed, and with the roles of the half-lines reversed.

Suppose there is an \( \alpha \) that ‘evenly’ divides \( P(\sigma, I) \), in the sense that
\[ P(\sigma 1_{(a,\infty)}, I), P(\sigma 1_{(-\infty,a)}, I) \geq \frac{1}{8} P(\sigma, I). \]
It then follows immediately from (3.4) that the \( L^2 \) product is controlled by \( N^2 \). (It is worth remarking that the role of the point \( \alpha \) is mysterious, and can be located far away from \( I \), see §11.)

Thus, we can assume that there is no such point \( \alpha \). That means that \( P(\sigma, I) \) is dominated by a point mass. For some \( \alpha \in \mathbb{R} \), there holds \( p_1^2(\sigma(\alpha)) \geq \frac{2}{3} P(\sigma, I) \). But then, \( w \) cannot have a point mass at \( \alpha \), whence we have, say, \( P(w 1_{(a,\infty)}, I) \geq \frac{1}{8} P(w, I) \), for some choice of \( \alpha > 0 \). Then, (3.4), with the roles of \( \sigma \) and \( w \) interchanged, and a slightly different choice of \( \alpha \) and \( \alpha \) proves the \( L^2 \) bound.

\[ \square \]

### 3.2. Choice of Truncations
Below, we will be analyzing ‘off-diagonal’ inner products, for which the ‘hard truncation’ definition of truncation in (1.2) is poorly suited. Let us agree to change the definition of the truncation. We begin with a very general truncation, and then pass to a choice of truncations that will minimize certain (well-known) technicalities.

For choices of constants \( 0 < \alpha < \beta \), take a function \( \psi_{\alpha,\beta}(y-x) \) that satisfies
\[ (\alpha + |y-x|)|\psi_{\alpha,\beta}(y-x)| \leq C 1_{|y-x|<2\beta}, \]
where \( C \) is absolute. Moreover, we have
\[ \psi_{\alpha,\beta}(y-x) = \frac{1}{y-x}, \quad \alpha < |x-y| < \beta. \]
Then, set the ‘smooth’ truncations of the Hilbert transform to be
\[ \tilde{H}_{\alpha,\beta} v(x) := \int \psi_{\alpha,\beta}(y-x) v(dy) \]
(3.5) \[ \frac{y}{y-x}. \]
(We will refer to them as ‘smooth’ as that is our goal. No gradient condition has been imposed on \( \psi_{\alpha,\beta}(y-x) \) just yet.)

There is another point about the truncation levels that is basic, and will be referred to below. Let \( f \) and \( g \) be two functions with
\[ 2\alpha < \inf_{x \in \text{supp}(f)} \inf_{y \in \text{supp}(g)} |x-y| \leq \sup_{x \in \text{supp}(f)} \inf_{y \in \text{supp}(g)} |x-y| < \beta, \]
(3.6)
then the inner product

\[ \langle \tilde{H}_{\alpha,\beta}(\sigma f), g \rangle_w = \iint f(y)g(x) \frac{\sigma(dy)}{y-x} \cdot w(dy) \cdot w(dy). \]  

(3.7)

The right side is independent of \( \alpha \) and \( \beta \). And this we take to be the canonical value of \( \langle H_{\sigma} f, g \rangle_w \), denoted with a superscript \( c \). It is noteworthy that a definition of canonical value also makes perfect sense for the 'hard' truncations, and only this property that enters into the proof of the \( A_2 \) bound.

The two types of truncations, 'hard' and 'smooth' are equivalent. We hence forth restrict attention to the 'smooth' truncations \( \tilde{H}_{\alpha,\beta} \).

**Proposition 3.8.** Let \( \sigma \) and \( w \) be two weights with no common point mass. One has the norm inequality

\[ \sup_{0<\alpha<\beta} \| H_{\alpha,\beta}(\sigma f) \|_w \leq \| f \|_{\sigma} \]

if and only if

\[ \sup_{0<\alpha<\beta} \| \tilde{H}_{\alpha,\beta}(\sigma f) \|_w \leq \| f \|_{\sigma}. \]

**Proof.** Let us see that the 'hard' truncations imply the 'smooth' truncations. The previous section shows that the \( A_2 \) bound is a consequence of the assumed norm inequality. Now, the difference \( H_{\alpha,\beta} - \tilde{H}_{\alpha,\beta} \) is controlled by the sum of two positive operators \( A_{\alpha} + A_{\beta} \), where

\[ A_{\alpha} f(x) = \frac{1}{4\alpha} \int_{x-2\alpha}^{x+2\alpha} f(y) \cdot dy. \]

(3.9)

This is the direct consequence of (3.5). And, the (simple) \( A_2 \) condition shows that \( A_{\alpha} \) satisfies the two weight inequality, independent of \( \alpha > 0 \), as we now show.

The argument is well-known. Calculate, for any \( \alpha > 0 \)

\[ \| A_{\alpha/2}(\sigma f) \|_w^2 = \int_{-\infty}^{\infty} \left[ \frac{1}{2\alpha} \int_{x-\alpha}^{x+\alpha} f(y) \cdot \sigma(dy) \right]^2 \cdot w(dx) \]

\[ \leq \int_{-\infty}^{\infty} \frac{1}{\alpha^2} \int_{x-\alpha}^{x+\alpha} f(y)^2 \cdot \sigma(dy) \cdot \sigma(x-\alpha, x+\alpha) \cdot w(dx) \]

\[ = \int_{-\infty}^{\infty} f(y)^2 \cdot \int_{y-\alpha}^{y+\alpha} \sigma(x-\alpha, x+\alpha) \cdot w(dx) \cdot \sigma(dy) \]

\[ \leq \int_{-\infty}^{\infty} f(y)^2 \cdot \frac{\sigma(y-2\alpha, y+2\alpha) \cdot w(y-\alpha, y+\alpha)}{\alpha^2} \cdot \sigma(dy) \leq A_2 \| f \|_{\sigma}^2. \]

Here, we have used Cauchy–Schwarz, and Fubini, to get to a point where the simple \( A_2 \) ratio reveals itself.

Conversely, one should check that if the smooth truncations are uniformly bounded, then the \( A_2 \) condition holds. But the proof of the \( A_2 \) bound only depends upon the canonical value of the truncations, as specified in (3.7), and this agrees for both the 'hard' and the 'smooth' truncations. Thus, the assumption of uniform norm bounds on the smooth truncations implies the \( A_2 \) bound. The averaging operators \( A_{\alpha} \) are then uniformly bounded, and one can pass to the 'hard' truncations.

\[ \square \]
It follows from this argument that the statement of our main theorem is insensitive to the choice of truncations: Provided it implies even the simple $A_2$ condition, then a wide family of truncations are bounded if and only if the original are bounded. (For much more on choices of truncations, see [28].)

We now make a particular choice of truncation. Consider a truncation given by

$$H_{\alpha,\beta}^{(s)}(f)(x) := \int f(y) K_{\alpha,\beta}(y-x) \sigma(dy)$$

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 $y > 0$, it is the minimal convex minorant of $1/x$ that agrees with $1/x$ on the interval $[\alpha, \beta]$. To be specific,

$$K_{\alpha,\beta}(y) := \begin{cases} -\frac{y}{\alpha} + \frac{2}{\alpha} & 0 < y < \alpha, \\ \frac{y}{\beta} - \frac{2}{\beta} & \beta < y < 2\beta, \\ 0 & 2\beta \leq y. \end{cases}$$  (3.10)

This is a $C^1$ function on $(0,2\beta)$ and is supported on $[0,2\beta]$. See Figure 5.

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

**Proposition 3.11.** Suppose that $2|x - x'| < |x - y|$, then

$$K_{\alpha,\beta}(y-x') - K_{\alpha,\beta}(y-x) = c_{x,x',y} \frac{x' - x}{(y-x)(y-x')},$$  (3.12)

where $c_{x,x',y} = 1$ if $2\alpha < |x-y| < \frac{2}{3} \beta$, and is otherwise positive and never more than one.

**Proof.** The assumptions imply that $y - x'$ and $y - x$ have the same sign. Assume, without loss of generality that $0 < y-x' < y-x$. It follows from the definition of $K_{\alpha,\beta}$ and the strict concavity of $1/y$
height = \frac{1}{|J|} P(\mu, J)

\text{Figure 6. An illustration of the monotonicity principle.}

Moreover if \(2\alpha < |x - y| < \frac{2}{3} \beta\), it follows that \(\alpha < |x' - y| < \beta\), whence equality holds on the right. \(\square\)

3.3. The Monotonicity Principle. Certain kinds of off-diagonal estimates for the Hilbert transform have concrete estimates in terms of the Poisson integral. This estimate makes this precise, and shows moreover that we need not be that careful about exactly which function appears in the Poisson integral. It is at the core of the entire proof.

Lemma 3.13 (Monotonicity Principle). Suppose that the two weights \(\sigma\) and \(w\) satisfy the \(A_2\) bound, and neither has a point mass at an endpoint of \(I\). Let \(J \subset I\). There holds for any \(g \in L^2(J, w)\), with \(w\)-integral zero,

\[
(3.14) \quad P(\sigma(\mathbb{R} - I), I) \langle \frac{x}{|I|}, \overline{\mathbf{g}} \rangle_w \leq \liminf_{\alpha \downarrow 0} \liminf_{\beta \uparrow \infty} \langle H_{\alpha, \beta}(\sigma(\mathbb{R} - I)), \mathbf{g} \rangle_w .
\]

Here, \(\overline{\mathbf{g}} = \sum_J |\mathbf{g}(J')| h^\mathbf{J}_w\), is a Haar multiplier applied to \(g\). Suppose that \(J \subset I\) is good, with \(2r|J| \leq |I|\). Then for any two compactly supported weights \(|\nu| \leq \mu\) supported off of the interval \(I\), there holds

\[
(3.15) \quad \sup_{0 < \alpha < \beta} \langle \langle H_{\alpha, \beta} \nu, g \rangle \rangle_w \leq \langle \langle H^\mu \nu, \overline{\mathbf{g}} \rangle \rangle_w \simeq P(\mu, J) \langle \frac{x}{|J|}, \overline{\mathbf{g}} \rangle_w .
\]

On the right, we use the canonical value of the inner product, as defined in (3.7).

Note that in the first estimate, the Poisson term is always estimated above by an inner product involving the Hilbert transform. In the second, note that the inner product can always be made larger by making the weight positive. Moreover, under moderate assumptions on the support of the weight, the first inequality can be reversed. See Figure 6 In that figure, the function \(\mu\) is outside of \(2^{[1-\epsilon]} J\), so that \(H\mu\) is a smooth increasing function on \(J\). Moreover, the derivative of \(H\mu\) is approximately \(|J|^{-1} P(\mu, J)\). So, if we form an inner product with the Haar function \(h^\mathbf{J}_w\), we only need to be concerned with the linear approximation to \(H\mu\). However, the conditions to get the reversal are particular, and this drives the case analysis in different sections of the proof.

Proof. We consider the first estimate. By linearity, it suffices to consider the case of \(g(x) = h^\mathbf{J}_w(x)\), for \(J \subset I\), and indeed we can take \(J = I\). We need to separate the two weights involved, so that we can pass to a canonical value of the inner product on the right in (3.14). The \(A_2\) condition is the only condition
needed for the weak-boundedness principle, Proposition 6.7. Applying it in this setting, notice that it shows that for \( \lambda > 1 \),

\[
|\langle H_\sigma(\lambda I - I), h^w_\sigma \rangle| \leq A_2^{1/2} \sqrt{\sigma(\lambda I - I)}.
\]

The assumption that \( \sigma \) does not have mass at the endpoints of \( I \) implies that \( \sigma(\lambda I - I) \) can be made arbitrarily small, as \( \lambda \downarrow 1 \). Therefore, it suffices to consider \( H_\sigma(\mathbb{R} - \lambda I) \), for some fixed \( \lambda > 1 \). Now, in addition, as \( \Lambda \uparrow \infty \), we have \( P(\sigma(\mathbb{R} - \lambda I), I) \to 0 \).

So, for \( 1 < \lambda < \Lambda < \infty \) fixed, the canonical version of the inner product, as given in (3.7), is

\[
\langle H^e(\sigma \cdot (\lambda I - \lambda I)), h^w_\lambda \rangle_w = \int_{\Lambda - \lambda I} \int_J \frac{1}{y-x} h^w_\lambda(x) w(dx) \sigma(dy).
\]

Here, \( x_\lambda \) is the center of \( J \), and it can be inserted for the usual reason that \( h^w_\lambda \) has \( w \)-integral zero. Then, use the fact that \( (x - x_\lambda) h^w_\lambda \geq 0 \), and that \( (y - x)(y - x_\lambda) > 0 \). So (3.14) holds.

The second inequality comes with the assumption that \( J \subset I \), \( 2^{|J|} < |I| \), whence \( \text{dist}(J, I) > |J|^{1/|I|^{1-\varepsilon}} \geq 2^{1-\varepsilon} |J| \). Namely, the support of \( h^w_\lambda \) and that of \( \mu \) are separated. Now, on the one hand, regardless of how the truncation levels \( \alpha, \beta \) are chosen, form the difference in the usual way.

\[
\langle H_{\alpha,\beta} \nu, h^w_\lambda \rangle_w = \int_{\mathbb{R} - I} \int_J \{K_{\alpha,\beta}(y - x) - K_{\alpha,\beta}(y - x_\lambda)\} h^w_\lambda(x) \nu(dy) w(dx).
\]

Here, \( x_\lambda \) is the center of \( J \). The assumptions on \( I \) and \( J \) imply that (3.12) applies to the difference on the kernel on the right. This gives us

\[
(3.16) \quad \int_{\mathbb{R} - I} \int_{J \times J} C_{x, y, J} \frac{x - x_\lambda}{y - x_\lambda} h^w_\lambda(x) \nu(dy) w(dx) \leq P(\mu, J) \cdot \langle \frac{x - x_\lambda}{|J|}, h^w_\lambda \rangle_w.
\]

This follows since \( C_{x, y, J} h^w_\lambda \geq 0 \), and \( |\nu| \leq \mu \).

To recap, we have shown that for all choices of truncation parameters, there holds

\[
\sup_{\alpha < \beta} \langle H_{\alpha,\beta} \nu, h^w_\lambda \rangle_w \leq P(\mu, J) \langle \frac{x}{|J|}, h^w_\lambda \rangle_w.
\]

It remains to show that for the positive weight \( \mu \),

\[
\langle H^e \mu, h^w_\lambda \rangle_w \simeq P(\mu, J) \langle \frac{x}{|J|}, h^w_\lambda \rangle_w.
\]

Recall that \( \mu \) and \( h^w_\lambda \) have separated support, in the sense required to define the canonical value of the inner product above, in the sense of (3.7). But, the canonical value is realized for particular values of \( \alpha \) and \( \beta \). Repeat the argument that lead to (3.16), but assuming that \( \nu = \mu \). Note that the constant \( C_{x, y, J} = 1 \), so the left side of (3.16) is positive, and the two sides are comparable by inspection. \( \square \)
3.4. The Energy Inequality. The energy inequality is phrased in terms of the quantity
\[ E(w, I)^2 := |I|^{-2} \langle w - I_w x \rangle^2 = |I|^{-2} \sum_{J : I \subset J} \langle x, h^{\alpha J} \rangle_w^2. \]

Lemma 3.17. [The Energy Inequality] For any interval \( I_0 \) and any partition \( P \) of \( I_0 \) into intervals such that neither \( \sigma \) nor \( w \) have point masses at the endpoints, there holds
\[ \sum_{I \in P} P(\sigma, I)^2 E(w, I)^2 w(I) \leq C_0 \mathcal{H}^2 \sigma(I_0). \]

Here, \( C_0 \) is an absolute constant.

Proof. It suffices to prove the inequality for all finite collections \( P \), with constant that is independent of the cardinality of \( P \). The fundamental fact is (3.14). It with simple duality considerations, and linearity of \( H_\sigma \) show that for each interval \( I \), for \( \alpha \) sufficiently small, and \( \beta \) sufficiently large,  
\[ \sum_{I \in P} \left\| H_{\alpha, \beta} \sigma(I_0 - I) \right\|_{L^2(I, w)}^2 \leq \left\| H_{\alpha, \beta} \sigma I_0 \right\|_{L^2(I, w)}^2 + \left\| H_{\alpha, \beta} \sigma I \right\|^2_{L^2(I, w)}. \]

These two terms are controlled by interval testing. The collection \( P \) is finite, so that fixed choices of \( \alpha \) and \( \beta \) are permitted, and so we have
\[
\begin{align*}
\sum_{I \in P} \left\| H_{\alpha, \beta} \sigma I \right\|_{L^2(I, w)}^2 &\leq \left\| H_{\alpha, \beta} \sigma I_0 \right\|_{L^2(I_0, w)}^2, \\
\sum_{I \in P} \left\| H_{\alpha, \beta} \sigma I \right\|_{L^2(I, w)}^2 &\leq \mathcal{H}^2 \sum_{I \in P} \sigma(I) \leq \mathcal{H}^2 \sigma(I_0).
\end{align*}
\]

Thus, it follows that we have this modified version of the energy inequality, in which the support of \( \sigma \) has ‘holes.’
\[ \sum_{I \in P} P(\sigma(I_0 - I), I)^2 E(w, I)^2 w(I) \leq C_0 \mathcal{H}^2 \sigma(I_0). \]

To get the Lemma as written, use the \( A_2 \) condition to ‘fill in the holes.’ \( \square \)

3.5. Context and Discussion.

3.5.1. The necessity of the \( A_2 \) condition was easily available, with an argument of Sergei Treil already pointed out by Sarason in his note [48]. This argument, based upon complex variables, has close analogs in [35, 58]. The real variable proof presented herein is in [20]. The early paper of Muckenhoupt and Wheeden [30] contains a proof of the necessity of the half-Poisson condition,
\[ \sup_I \frac{\sigma(I)}{|I|} P(w, I) + \frac{w(I)}{|I|} P(\sigma, I) \leq N^2. \]

(The half-Poisson condition suffices for almost the entire proof; the full Poisson condition is needed to deduce the functional energy inequality.) Higher dimensional extensions, which are not straightforward, are discussed in [19].
3.5.2. The energy inequality was influenced by the following assumption placed upon the pair of weights in [35, 58]. Assume that there is a finite constant \( P \) so that for all intervals \( I_0 \), and all partitions \( P \) of \( I_0 \) into intervals,
\[
\sum_{I \in P} P(\sigma \cdot I_0, I)^2 w(I) \leq P^2 \sigma(I_0).
\]
(3.19)

Also assume that the dual inequality holds. In the language of Nazarov-Treil-Volberg, this is the **pivotal condition**. They proved

**Theorem C.** Assume that \( w \) and \( \sigma \) do not share a common point mass. Then, there holds \( N \leq A_2^{1/2} + T + P \).

This is a very strong Theorem, with an important proof. It decisively used the tools of non-homogeneous harmonic analysis, namely random grids, good-bad projections. The pivotal condition controlled certain degeneracies in the pair of weights, compare to Definition 4.5. To illustrate the difficulties in the general case, we prove this theorem in §10.

The pivotal condition holds if the pair of maximal function estimates hold, namely \( M_\sigma : L^2(\sigma) \rightarrow L^2(w) \) and \( M_w : L^2(w) \rightarrow L^2(\sigma) \). This is easy to see. From (3.19),
\[
\sum_{I \in P} P(\sigma \cdot I_0, I)^2 w(I) \leq \sum_{I \in P} \inf_{x \in I} M(\sigma \cdot I_0) (x)^2 w(I)
\]
\[
\leq \int_{I_0} M(\sigma \cdot I_0)^2 \, dw \leq \sigma(I_0),
\]
by the assumed norm bound on the maximal function. One sees that Theorem C offered a complete characterization of the two weight inequality for the triple of operators \( (H_\sigma, M_\sigma, M_w) \). If the pair of weights are doubling, then the boundedness of the maximal functions is a consequence of the \( A_2 \) condition.\(^4\) The full characterization of the boundedness of the Hilbert transform was thus known for doubling measures. See [58].

The pivotal condition is generic in the following sense. Assuming the pivotal condition, the Hilbert transform can be replaced by a generic Calderón-Zygmund operator with one derivative on its kernel. This, and its extension to operators with a rougher kernel, was fundamental in the paper [42], whose main result was an important intermediate one in the solution of the \( A_2 \) conjecture [12].

3.6. The functional energy inequality is also generic, in the following sense, which we describe a little imprecisely. Given a pair of weights \( \sigma, w \) let \( \mathcal{E} \) be the best constant in the inequality (7.3), as written, and in its dual form. Let \( T_T f(x) = \int K(x,y) f(y) \, dy \) be an whose kernel \( K(y) \) satisfies the size and gradient condition
\[
|x - y| : |\nabla K(x,y)| + |K(x,y)| \leq |x - y|^{-1}.
\]

Let \( N_T \) be the norm of \( T_\sigma \) from \( L^2(\sigma) \) to \( L^2(w) \), and let \( \mathcal{T} \) be the best constant in the testing inequalities
\[
\int_I |T_\sigma 1_I|^2 \, dw \leq \mathcal{T}^2 \sigma(I), \quad \text{and} \quad \int_I |T_w 1_I|^2 \, d\sigma \leq \mathcal{T}^2 w(I),
\]
Then, there holds \( N_T \leq A_2^{1/2} + \mathcal{E} + \mathcal{T} \). The paper [52] formalizes this, for a class of fractional Calderón-Zygmund operators.

\(^4\) Alternatively, under the assumption of \( w \) being doubling, check that the energy satisfies \( \mathcal{E}(w, I) \geq 1 \), with the implied constant depending upon the doubling constant. Thus, the necessary energy inequality implies the pivotal condition.
Figure 7. The function \( \frac{|I|}{|I|^2 + |x-x_I|^2} \) are graphed for three separate intervals.

Observe that \( T \) need not be bounded on \( L^2(dx) \), and that requiring one derivative on the kernel is important. If fewer derivatives are required, then the Poisson integral needs to be changed to reflect the rougher kernel.

3.6.1. Nazarov-Treil-Volberg, in language reminiscent of Sarason, wrote that ‘perhaps the pivotal condition is necessary’ for the boundedness of the Hilbert transform. This turned out to have a strong measure of truth, in that using the specific structure of the Hilbert transform, the energy inequality was shown necessary in [20]. Note that one can formally obtain the pivotal condition (3.19) from the energy inequality (3.18) by raising the energy term \( E(w, I) \) to the zero power, rather than the necessary power 2. The paper [20] then adapted the approach of [35, 58], essentially imposing a new weaker condition on the pair of weights in which one raised the energy to a power intermediate between 0 and 2. In addition, that paper provided an explicit example, recounted in §11, that showed that the pivotal condition (3.19) is not necessary for the boundedness of the Hilbert transform.

3.6.2. The energy inequality is rather subtle. The Poisson term \( P(\sigma, I) \) can be much larger than the simple average, but this is compensated for with the terms \( E(w, I)^2 w(I) \). The Figure 7 is offered to provide some insight into the ‘long tails’ that the Poisson term can have.

Another indication of this subtlety is the observation that the energy inequality will not follow from just the \( A_2 \) condition. Given interval \( I_0 \), and partition \( P \) of \( I_0 \), one can write

\[
\sum_{I \in P} P(\sigma, I)^2 E(w, I)^2 w(I) \leq A_2 \sum_{I \in P} |I| \cdot P(\sigma, I)^2
\]

\[
= A_2 \int_{I_0} \sum_{I \in P} \frac{|I|^2}{(|I| + \text{dist}(x, I))^2} \sigma(dx).
\]

To finish, one would have to know that the function inside the integral is bounded. But, this is not true in general. Though a very tame BMO function, this fact does not help, since \( \sigma \) is a general measure, and need not satisfy any \( A_{\infty} \) type condition. Indeed, the proof of the main theorem would be more or less classical if the weights satisfy a \( A_{\infty} \) type conditions.

3.6.3. The monotonicity principle, Lemma 3.13, was noted in [21]. It, with the energy inequality, are essential aspects of the proof.
3.6.4. If the pair of weights share a common point mass, this presents no difficulty in interpretation of the norm estimate as uniform over all truncation. But, the $A_2$ condition (1.4) will fail if the pair of weights share a common point mass. In turn, this implies that the averaging operators (3.9) need not be uniformly bounded: The formulation of the norm inequality could depend critically on the choice of truncation operator.

Our main theorem will hold for weights that are supported on uniformly sparse sets, like $\mathbb{Z}$, as a subset of $\mathbb{R}$, this is because the scales of length less than one are negligible, and a choice of truncation will be irrelevant to the result. But, if the weights are, say, supported on point masses at the rational points, one’s difficulties seem enormous.

4. Global to Local Reduction

Our aim is to prove the estimate (2.5),

$$\sup_{\alpha, \beta} |\langle H_{\alpha, \beta} (\sigma P_\text{good}^w f), P_\text{good}^w g \rangle_w| \lesssim \varepsilon \| f \|_w \| g \|_w.$$

That is, the bilinear form only needs to be controlled for $(\varepsilon, r)$-good functions $f = P_\text{good}^w f$ and similarly for $g$, goodness being defined with respect to a fixed dyadic grid. Suppressing the notation, we write “good” for “$(\varepsilon, r)$-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 $g$. We clearly remark on goodness when the property is used; any value of $0 < \varepsilon \leq \frac{1}{4}$ is sufficient for our purposes. The symbol $\varepsilon$ is kept throughout, as a guide to the appearance of the good property of intervals.

The notation $H_{\alpha, \beta}$ is for the smooth truncations given by the kernel in (3.10). We further suppress the notation, writing simply $\langle H_{\sigma} f, g \rangle_w$. The inequality above is reduced to the local estimate, (4.17), at the end of this section. 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 $h_{I_0}^f$ for $I \subset I_0$, and similarly for $g$.

The distinction between $J \subset I$ and $J \preceq I$ ($J \subset I$ and $2^r|J| \leq |I|$) forces some case analysis. This is further simplified by this assumption on the Haar supports of $f, g$. There are two integers $s_f, s_g$ such that

$$f = \sum_{I: I \subset I_0} \Delta_I^f$$

and similarly for $g$. Thus, the lengths of the (good) intervals $I$ are restricted to an equivalence class mod $r$. Set $\mathcal{D}_f := \{ I : \log_2|I| \in s_f - 1 + r\mathbb{Z} \}$, so these are the children of the intervals that appear in (4.1).

Then, we are to control the bilinear form

$$\langle H_{\sigma} f, g \rangle_w = \sum_{I, J: I \preceq J \subset I_0} \langle H_{\sigma} \Delta_I^f, \Delta_J^w g \rangle_w.$$

The sum is broken into many summands. The most important of these are the two ‘triangular’ forms

$$B^{\text{above}}(f, g) := \sum_{I: I \subset I_0} \sum_{J \preceq I} E_{I, J}^w \Delta_I^f \cdot \langle H_{\sigma} I_\sigma, \Delta_J^w g \rangle_w$$

and the dual form, $B^{\text{below}}(f, g)$. Here, $J \preceq I$ means that $J \subset I$ and $2^r|J| \leq |J|$, in words ‘$J$ is strongly contained in $I$’. Goodness of $J$ justifies the use of this condition. A basic fact, proved in §6, is
Lemma 4.3. There holds

\[ |\langle H_\sigma f, g \rangle_w - B_{\text{above}}(f, g) - B_{\text{below}}(f, g) \rangle \leq \mathcal{H}\|f\|_\sigma \|g\|_w. \]

Thus, the main technical result is as below; it immediately supplies our main theorem.

Theorem 4.4. There holds

\[ |B_{\text{above}}(f, g) \rangle \leq \mathcal{H}\|f\|_\sigma \|g\|_w. \]

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

In the remainder of this section is devoted to a reduction of the global Theorem 4.4 to a local estimate described in Theorem 4.11. In the local estimate, the function \( f \) is more structured in that it has bounded averages on a fixed interval, and the pair of functions \( f, g \) are more structured in that their Haar supports avoid intervals that strongly violate the energy inequality, in the following sense.

The decomposition of the triangular form begins with the selection of stopping data. We construct stopping data, which accomplishes two ends, in that it will control certain telescoping sums of martingale differences of \( f \), and that it controls certain degeneracies in an energy estimate on the weights.

Definition 4.5. Given any interval \( I_0 \), define \( F_{\text{energy}}(I_0) \) to be the maximal subintervals \( I \subset I_0 \) such that

\[ \mathcal{P}(\sigma : I_0, I)^2 E(w, I)^2 w(I) > 10C_0 \mathcal{H}^2 \sigma(I). \]

Here, \( C_0 \) is the constant in (3.18). There holds \( \sigma(\cup \{ F : F \in F(I_0) \}) \leq \frac{1}{10} \sigma(I_0) \), by the energy inequality.

We make the following construction of stopping intervals \( F \). Add \( I_0 \) to \( F \), and set \( \alpha_f(I_0) := E_\sigma f(I_0) \).

In the inductive stage, if \( F \in F \) is minimal, add to \( F \) those maximal descendants \( F' \in D_f \) of \( F \) such that either (a) \( E_{F'} f \geq 10 \alpha_f(F) \), or (b) \( F' \in F_{\text{energy}}(F) \). Then define

\[ \alpha_f(F') := \begin{cases} \alpha_f(F) & \text{if } E_{F'} f < 2\alpha_f(F) \\ \frac{E_{F'} f}{\sigma} & \text{otherwise} \end{cases}. \]

If there are no such intervals \( F' \), the construction stops. We refer to \( F \) and \( \alpha_f(\cdot) \) as Calderón-Zygmund stopping data for \( f \). Their key properties are collected here.

Lemma 4.6. For \( F \) and \( \alpha_f(\cdot) \) as defined above, there holds

1. \( I_0 \) is the maximal element of \( F \).
2. For all \( I \in D \), \( I \subset I_0 \), we have \( |E f| \leq 10 \alpha_f(\pi_f I) \).
3. \( \alpha_f \) is monotonic: If \( F, F' \in F \) and \( F \subset F' \) then \( \alpha_f(F) \geq \alpha_f(F') \).
4. The collection \( F \) is \( \sigma \)-Carleson in that

\[ \sum_{F \in F : F \subset S} \sigma(F) \leq 2 \sigma(S), \quad S \in D. \]

5. We have the inequality

\[ \sum_{F \in F} \alpha_f(F) \cdot F \leq \|f\|_\sigma. \]
Proof. The first three properties are immediate from the construction. The fourth, the \( \sigma \)-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

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

Or, they satisfy \( \mathbb{E}_{F}^{G}[|f|] \geq 10 \alpha_{r}(F) \geq 5 \mathbb{E}_{G}[|f|] \), as follows from the construction above. These intervals then satisfy a similar estimate. Hence, (4.7) holds.

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

\[
\Phi_{G} := \sum_{F \in \mathcal{F}: \pi_{F} = G} F.
\]

Define \( G_{k} := \{ \Phi_{G} \geq 2^{k} \} \), for \( k = 0, 1, \ldots \). The \( \sigma \)-Carleson property implies integrability of all orders in \( \sigma \)-measure of \( \Phi_{G} \). Using the third moment, we have \( \sigma(G_{k}) \leq 2^{-3k} \sigma(G) \). Namely, expanding the integral and using the Carleson measure property of \( \mathcal{F} \),

\[
\int_{\mathcal{F}} \Phi_{G}^{2} \sigma(dx) = \sum_{F_{1}, F_{2}, F_{3} \in \mathcal{F}} \sigma(F_{1} \cap F_{2} \cap F_{3}) \leq 6 \sum_{F_{1}, F_{2}, F_{3} \in \mathcal{F}} \sigma(F_{3}) \leq \sigma(G_{k}) \leq \sigma(G).
\]

It follows that \( \sigma(G_{k}) \leq 2^{-3k} \sigma(G) \).

Then, estimate

\[
\left\| \sum_{F \in \mathcal{F}} \alpha_{r}(F) \cdot F \right\|_{\sigma}^{2} = \left\| \sum_{G \in \mathcal{G}} \alpha_{r}(G)\Phi_{G} \right\|_{\sigma}^{2} \leq \sum_{k=0}^{\infty} (k + 1)^{1/2} \left\| \sum_{G \in \mathcal{G}} \alpha_{r}(G)2^{k} \mathbf{1}_{G_{k}} \right\|_{\sigma}^{2} \leq \sum_{k=0}^{\infty} (k + 1) \left\| \sum_{G \in \mathcal{G}} \alpha_{r}(G)2^{k} \mathbf{1}_{G_{k}}(x) \right\|_{\sigma}^{2} \leq \sum_{k=0}^{\infty} (k + 1)^{2} \left\| \sum_{G \in \mathcal{G}} \alpha_{r}(G)2^{2k} \sigma(G_{k}) \right\| \leq \sum_{G \in \mathcal{G}} \alpha_{r}(G)^{2} \sigma(G) \leq \|Mf\|_{\sigma}^{2} \leq \|f\|_{\sigma}^{2}.
\]

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

We will use the notation
\[
P_\sigma^F f := \sum_{I : \pi_F^{-1} = F} \Delta_I^f, \quad F \in \mathcal{F}.
\]
and similarly for \( Q_w^F g \). (Note that both are projections, but \( P_\sigma^F f \) is a structured function, while \( Q_w^F g \) is not.) The inequality (4.8) allows us to estimate
\[
\sum_{F \in \mathcal{F}} \{ \alpha_F(F) \sigma(F) \}^{1/2} \| P_\sigma^F f \|_{\sigma} \| Q_w^F g \|_w
\]
\[
\leq \left[ \sum_{F \in \mathcal{F}} \{ \alpha_F(F)^2 \sigma(F) + \| P_\sigma^F f \|_{\sigma}^2 \} \times \sum_{F \in \mathcal{F}} \| Q_w^F g \|_w^2 \right]^{1/2} \leq \| f \|_{\sigma} \| g \|_w.
\]
We will refer to as the quasi-orthogonality argument. It holds only under the assumption that the projections \( Q_w^F g \) are pairwise orthogonal. It is very useful.

Return to the double sum (4.2). Using \( \mathcal{F} \) as constructed, note that
\[
(4.10) \quad B_{\mathcal{F}, \text{loc}}^{\text{above}}(f, g) = B_{\mathcal{F}, \text{glob}}^{\text{above}}(f, g) + B_{\mathcal{F}, \text{glob}}^{\text{above}}(f, g),
\]
\[
B_{\mathcal{F}, \text{loc}}^{\text{above}}(f, g) := \sum_{F \in \mathcal{F}} \sum_{I : \pi_F^{-1} = F} \sum_{J : \pi_F^{-1} = F \setminus I} B_{I, J} \Delta_I^f \cdot \langle H_{\sigma} I, \Delta_J^g \rangle
\]
\[
= \sum_{F \in \mathcal{F}} B_{\mathcal{F}, \text{glob}}^{\text{above}}(P_\sigma^F f, Q_w^F g)
\]
\[
B_{\mathcal{F}, \text{glob}}^{\text{above}}(f, g) := \sum_{I, J : J \subseteq \pi_F^{-1} \setminus I \subseteq \pi_F^{-1}} B_{I, J} \Delta_I^f \cdot \langle H_{\sigma} I, \Delta_J^g \rangle
\]

We henceforth concentrate on the ‘above’ forms, with all considerations applying in their dual formulation to control the ‘below’ forms. The global to local reduction is:

**Corollary 4.11.** [Global to Local Reduction] There holds
\[
| B_{\mathcal{F}, \text{glob}}^{\text{above}}(f, g) | \leq \mathcal{O}(\| f \|_{\sigma} \| g \|_w).
\]

**Proof.** This is a corollary to our functional energy estimate Theorem 7.2. Hence, we need to translate the inequality above to one about the Poisson sums in Theorem 7.2. But, this can’t be done immediately, due to the difference between \( J \subseteq I \) and \( J \subset I \). It is this distinction that we address first.

We have the tools to address these three subcases.

A: \( J \subseteq \pi_F^{-1} \subset I \). (\( J \) is strongly contained its \( \mathcal{F} \)-parent.)

B: \( J \notin \pi_F^{-1} \subseteq \pi_F^{-1} \subset I \). (\( J \) is only strongly contained in it’s \( \mathcal{F} \)-grandparent, which is contained in \( I \).)

C: \( J \subseteq I \), \( J \notin \pi_F^{-1} \subset I \subseteq \pi_F^{-1} \). (\( J \) is not strongly contained in its \( \mathcal{F} \)-parent, and \( I \) is contained in the \( \mathcal{F} \)-grandparent.)

Here we are using the notation that immediately proceeds Definition 7.1, and the reduction to these three cases is given below.
Case A. The functional energy inequality is most easily applicable in this case. Set $g_F := \sum_{1 \leq F = \pi_J} J^\sigma g$. This class of functions is $\mathcal{F}_{w-\alpha}$-adapted in the sense of Definition 7.1. This is in fact the only property of the $g_F$ that is needed for the argument below, which point is relevant to Case B below.

The sum in question is

$$\sum_{F \in \mathcal{F}} \sum_{1 \leq I \leq 2F} E^\sigma_{I,F} \Delta_I^\sigma f \cdot \langle H_\sigma I_F, g_F \rangle_w.$$

We invoke, for the first time, the Hilbert-Poisson exchange argument: (a) Replace the argument of the Hilbert transform by a stopping interval. (b) Invoke the stopping tree construction to control the sum of martingale differences of $f$. (c) Apply interval testing, on the stopping interval. (d) Use the monotonicity principle to dominate the complementary term in terms of a Poisson integral. (e) Apply (functional) energy to control the Poisson term. (f) Use quasi-orthogonality, as needed.

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 $H_\sigma$. Note that by the standard martingale difference identity and the construction of stopping data,

$$\left| \sum_{1 \leq I \leq 2F} E^\sigma_{I,F} \Delta_I^\sigma f \right| \leq \alpha_f(F), \quad F \in \mathcal{F}.$$

Hence, invoking interval testing,

$$\left| \sum_{F \in \mathcal{F}} \sum_{1 \leq I \leq 2F} E^\sigma_{I,F} \Delta_I^\sigma f \cdot \langle H_\sigma I_F, g_F \rangle_w \right| \leq \sum_{F \in \mathcal{F}} \alpha_f(F) \left| \langle H_\sigma I_F, g_F \rangle_w \right| \leq 2\mathcal{F} \sum_{F \in \mathcal{F}} \alpha_f(F) \sigma(F)^{1/2} \| g_F \|_w.$$

Quasi-orthogonality bounds this last expression.

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

$$\left| \sum_{1 \leq I \leq 2F} E^\sigma_{I,F} \Delta_I^\sigma f \cdot (I_F - F) \right| \leq \Phi := \sum_{F' \in \mathcal{F}} \alpha_f(F') \cdot F', \quad F \in \mathcal{F}.$$

Therefore, by the definition of $\mathcal{F}_{w-\alpha}$-adapted, the monotonicity property (3.15) applies, and yields

$$\sum_{I \leq F} E^\sigma_{I,F} \Delta_I^\sigma f \cdot \langle H_\sigma (I_F - F), g_F \rangle_w \leq \sum_{J \in \mathcal{J}^w(F)} \Phi(J) \left( \frac{x}{J} \right)^{1/2} \| g_F \|_w, \quad F \in \mathcal{F}.$$

Here $\Phi := \sum_{J \in \mathcal{J}^w(F)} \Phi(J) \cdot h_J^w$, so that every term has a positive inner product with $x$, and $\mathcal{J}^w(F)$ are the maximal good intervals $I \subseteq F$.

The sum over $F \in \mathcal{F}$ of (4.12) is controlled by functional energy, and the property that $\| \Phi \|_\sigma \leq \| f \|_\sigma$. We have proved

$$\sum_{F \in \mathcal{F}} \sum_{1 \leq I \leq 2F} E^\sigma_{I,F} \Delta_I^\sigma f \cdot \langle H_\sigma I_F, g_F \rangle_w \leq 2\mathcal{F} \| f \|_\sigma \| g \|_w.$$

This completes the Hilbert-Poisson exchange argument.

Case B. There are consequences of the assumption (4.1), on the Haar supports of $f$ and $g$. As well, except for the root, $\mathcal{F}$ is a subset of $\mathcal{D}_r$, defined just after (4.1). Suppose $J \subseteq I$ but $J$ is not strongly contained in its $\mathcal{F}$-parent $F$. We can assume that the parent of $J$ is not the root $I_0$. Thus $F \in \mathcal{D}_r$. It follows that there is an integer $0 \leq t < r$ such that $2^t |J| = |F|$. And, $J \in \pi_{2^t} J$. Here we are using the
notation that precedes Definition 7.1, so that $\pi^{i}_{F}J = \pi^{i}_{F}F$ is the grandparent of $J$ in the $F$-tree. These are the basic combinatorial observations, which we combine with the previous argument to complete Case B and Case C.

We continue with Case B, defining

$$\tilde{g}_{F} := \sum_{F': \pi^{i}_{F}F' = F} \sum_{2^{i}|J|=|F'|} \Delta_{j}^{w}g.$$

Observe that this sequence of functions is $F_{-}$-adapted, and hence, by (4.13),

$$\left| \sum_{F \in F} \sum_{I \in I_{F}} E_{I_{j}}^{g_{F}} \Delta_{j}^{w}f \cdot \langle H_{\sigma}I_{F}, \tilde{g}_{F}\rangle_{w} \right| \leq J\|f\|_{\sigma}\|g\|_{w}.$$

**Case C.** We give an upper bound for the sums below.

$$\sum_{F \in F} \sum_{J \in J_{F}} \sum_{|I| = |F|} E_{I_{j}}^{g_{F}} \Delta_{j}^{w}f \cdot \langle H_{\sigma}I_{F}, \Delta_{j}^{w}g \rangle_{w}, \quad 0 \leq t < r.$$

Note that they exhaust Case C, since we are fixing the relative lengths of $J$ and its $F$-parent.

We will perform the Hilbert-Poisson exchange argument, replacing the intervals $I_{F}$ above by the larger interval $F$. Define $\varepsilon_{F}$ by

$$\varepsilon_{F} = \sum_{J \in J_{F}} \|E_{I_{j}}^{g} \Delta_{j}^{w}f \cdot \langle H_{\sigma}I_{F}, \Delta_{j}^{w}g \rangle_{w} \|_{w}.$$

Then, $|\varepsilon_{F}| \leq 1$, and we have

$$\left| \sum_{F: \pi^{i}_{F}F = F} \sum_{J : 2^{i}|J| = |F|} \sum_{I \in I_{F}} E_{I_{j}}^{g_{F}} \Delta_{j}^{w}f \cdot \langle H_{\sigma}I_{F}, \Delta_{j}^{w}g \rangle_{w} \right|$$

$$= \varepsilon_{F} \left| \sum_{F: \pi^{i}_{F}F = F} \sum_{J : 2^{i}|J| = |F|} \Delta_{j}^{w}g \right|_{w}$$

$$\leq \alpha_{F}(\hat{F}) \sigma(\hat{F})^{1/2} \left\| \sum_{F: \pi^{i}_{F}F = F} \sum_{J \in J_{F}} \Delta_{j}^{w}g \right\|_{w}.$$

(4.14)

And quasi-orthogonality completes this bound.

For the remainder, we use the monotonicity principle to see that for any $F$ with $\pi^{i}_{F}F = \hat{F}$,

$$\left| \sum_{I \in I_{F} \subset F} \sum_{J : 2^{i-1}|J| = |I|} E_{I_{j}}^{g_{F}} \Delta_{j}^{w}f \cdot \langle H_{\sigma}(\hat{F} - I_{F}), \Delta_{j}^{w}g \rangle_{w} \right|$$

$$\leq \alpha_{F}(\hat{F})P(\sigma \cdot \hat{F}, F)E(w, F)w(F) \left\| \sum_{J : 2^{i}|J| = |I|, J \in J_{F}} \Delta_{j}^{w}g \right\|_{w}.$$

To control the sum over $F$ with with $\pi^{i}_{F}F = \hat{F}$, use Cauchy–Schwarz and the energy inequality to get the upper bound in (4.14). Then, apply quasi-orthogonality to complete the proof of the global to local reduction. \[\square\]
It remains to control \( B_{\mathcal{F}, loc}^{\text{above}}(f, g) \). Keeping the quasi-orthogonality argument in mind, appropriate control on the individual summands is enough to control it. To describe what has been done, one must note that the functions \( P_{\mathcal{F}}r f \) need not be bounded. But, we are only concerned with averages over intervals where the average will be bounded. In addition this function and \( Q_{\mathcal{F}}^w g \) are well-adapted to the pair of weights \( w, \sigma \). This is formalized in the next definition.

**Definition 4.15.** Let \( I_0 \) be an interval, and let \( \mathcal{S} \) be a collection of disjoint intervals contained in \( S \). A function \( f \in L^2(I_0, \sigma) \) is said to be uniform \((w.r.t. \mathcal{S})\) if these 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 \( I \) which is not contained in any \( S \in \mathcal{S} \), \( |\mathbb{E}_s^f| \leq 1 \).

We will say that \( g \) is adapted to a function \( f \) uniform \((w.r.t. \mathcal{S})\), if \( g \) is constant on each interval \( S \in \mathcal{S} \). We will also say that \( g \) is adapted to \( \mathcal{S} \).

In this next Lemma, the hypothesis is the local estimate, and the conclusion is the required bound on \( B_{\mathcal{F}, loc}^{\text{above}}(f, g) \). Note that the local estimate is homogeneous in \( g \), but not \( f \), since the term \( \sigma(I_0)^{1/2} \) on the right is motivated by the bounded averages property of \( f \).

**Lemma 4.16.** \([\text{The Local Estimate}]\) Assume that

\[
|B_{\mathcal{F}, loc}^{\text{above}}(f, g)| \leq \mathcal{F}(\sigma(I_0)^{1/2} + ||f||_w) ||g||_w,
\]

where \( f, g \) are of mean zero on their respective spaces, supported on an interval \( I_0 \). Moreover, \( f \) is uniform, and \( g \) is adapted to \( f \). Then, there holds, for all \( f \) and \( g \),

\[
|B_{\mathcal{F}, loc}^{\text{above}}(f, g)| \leq \mathcal{F}(||f||_w ||g||_w).
\]

**Proof.** Let \( \mathcal{F} \) and \( \alpha_{\mathcal{F}}(\cdot) \) be standard Calderón-Zygmund stopping data for \( f \). By Lemma 4.3, it suffices to bound

\[
B_{\mathcal{F}, loc}^{\text{above}}(f, g) = \sum_{F \in \mathcal{F}} B_{F}^{\text{above}}(P_{\mathcal{F}}^w f, Q_{\mathcal{F}}^w g)
\]

Observe that for an absolute constant \( C \), the function

\[
(C\alpha_{\mathcal{F}}(F))^{-1} P_{\mathcal{F}}^r f
\]

is uniform on \( F \) w.r.t. \( \mathcal{S}_F \), the \( \mathcal{F} \)-children of \( F \). Moreover, the function \( Q_{\mathcal{F}}^w g \) does not have any interval \( J \) in its Haar support contained in an interval \( S \in \mathcal{S}_F \). That is, it is adapted to the function in (4.18). Therefore, by assumption,

\[
|B_{\mathcal{F}}^{\text{above}}(P_{\mathcal{F}}^w f, Q_{\mathcal{F}}^w g)| \leq \mathcal{F}(\alpha_{\mathcal{F}}(F) \sigma(F))^{1/2} + ||P_{\mathcal{F}}^w f||_w ||Q_{\mathcal{F}}^w g||_w.
\]

The sum over \( F \in \mathcal{F} \) of the right hand side is bounded by the quasi-orthogonality argument of (4.9). \( \Box \)

Thus, it remains to show that the local estimate (4.17) holds. The Hilbert-Poisson exchange argument is invoked, but we will arrive at a Poisson term which falls outside the immediate scope of the energy inequality. Focusing on the argument of the Hilbert transform in (4.17), we write \( I_1 = I_0 - (I_0 - I_1) \). When the interval is \( I_0 \), and \( J \) is in the Haar support of \( g \), notice that the scalar

\[
\epsilon_j := \sum_{I: J \subseteq I \subseteq I_0} \mathbb{E}_I^\sigma \Delta_j^\sigma f
\]
is bounded by one: Say that $f$ is uniform w.r.t. $S$, and let $I$ be the minimal interval in the Haar support of $f$ with $J \subset I$. Since $g$ is adapted to $f$, we cannot have $I$ contained in an interval of $S$, and so $|E^I f| \leq 1$. By the telescoping identity for martingale differences,

$$
\varepsilon_I = \sum_{1: 1 \subset I \subset I_0} E^I f \cdot \langle H_{\sigma I_0}, \Delta_w^I g \rangle_w = \sum_{J: J \subset I_0} \varepsilon_J \Delta_w^J g \cdot \langle H_{\sigma I_0}, \Delta_w^J g \rangle_w
$$

which is at most one in absolute value.

Therefore, we can write

$$
\left| \sum_{1: I \subset I_0} \sum_{J: J \subset I} E^J f \cdot \langle H_{\sigma I_0}, \Delta_w^J g \rangle_w = \left| \langle H_{\sigma I_0}, \sum_{J: J \subset I_0} \varepsilon_J \Delta_w^J g \rangle_w \right|
\right| \leq \mathcal{T}_\sigma(I_0)^{1/2} \left| \sum_{J: J \subset I_0} \varepsilon_J \Delta_w^J g \right|_w \leq \mathcal{T}_\sigma(I_0)^{1/2} \| g \|_w.
$$

(4.19)

This uses only interval testing and orthogonality of the martingale differences, and it matches the first half of the right hand side of (4.17).

When the argument of the Hilbert transform is $I_0 - I_J$, this is the stopping form, the last component of the local part of the problem. It requires a subtle recursion, described in §5.

4.1. Context and Discussion.

4.1.1. Many T1 theorems have arguments, sometimes subtle ones, about telescoping sums which collapse. These arguments are systematically handled herein with the stopping data, as opposed to more intricate Carleson measure arguments.

4.1.2. The use of the energy stopping intervals, Definition 4.5, is motivated by the use of the corresponding intervals, under the pivotal condition (3.19), in [35, 58]. However, the pivotal condition is not necessary for the two weight inequality, while the energy inequality is necessary from the $A_2$ and interval testing conditions.

4.1.3. Initial arguments had largely ignored the structure of the pair of functions $f, g$ in the inner product $\langle H_{\sigma f}, g \rangle_w$, instead concentrating on proving an intricate series of Carleson measure type estimates. This changed with the argument of [21], which introduced Calderón-Zygmund stopping intervals, and the quasi-orthogonality argument into the subject. It was only then that the functional energy inequality was identified, but not proved, in [21]. Stopping data also allows us to avoid the subtle problem of absence of canonical paraproducts. Attempts to introduce them induce ad hoc elements into the proof.

4.1.4. The use of the functional energy inequality in this context follows the argument of [22], where this inequality was proved.

4.1.5. This section begins with the elementary and familiar Lemma 4.3, and then argues that the control of the triangular form $B^{\text{above}}(f, g)$ splits into the 'global to local' and the 'local' part. The authors of [22] only had the first reduction. And, using the techniques of that paper, could prove
Theorem D. ([22]) There holds \(|B_{\text{above}}(f, g)| \leq (\mathcal{H} + B_{\infty})||f||_w||g||_w||w\), where \(\mathcal{H} = A_{2}^{1/2} + \mathcal{I}\), and the remaining constant is the best constant in

\[|B_{\text{above}}(f, g)| \leq B_{\infty} \sigma(I_0)^{1/2}||g||_w,\]

where \(|f| \leq 1_{I_0}\), and \(I_0\) is any interval. The corresponding estimate holds for the dual from \(B_{\text{below}}(f, g)\).

This is a powerful Theorem, strongly suggesting that the \(A_2\) condition and testing the Hilbert transform over bounded functions is sufficient for the \(L^2\) boundedness of \(H_\sigma\). But, there is no obvious way to deduce such a result from the Theorem above. Phrasing things differently, it can be very difficult to translate partial information about the triangular form \(B_{\text{above}}(f, g)\) to information about \(\langle H_\sigma f, g \rangle_w\), a potentially serious obstacle if a richer theory of two weight inequalities for singular integrals is to be developed.

The parallel corona was introduced in [23] to surmount this obstacle. With it, the result that could be proved the first real variable characterization of the two weight inequality for any continuous singular integral.

Theorem E. [Lacey Sawyer Shen Uriarte-Tuero [23]] There holds \(N \simeq A_{2}^{1/2} + \mathcal{I}_{\infty}\), where the latter constant is the best constant in the inequalities below, uniform over all intervals \(I\), and Borel subsets \(E \subset I\).

\[\int_I |H_\sigma 1_E|^2 \, d\sigma \leq \mathcal{I}_{\infty}^2 \sigma(I), \quad \int_I |H_\sigma 1_E|^2 \, d\sigma \leq \mathcal{I}_{\infty}^2 w(I).\]

(One tests the Hilbert transform on \(1_E\), but only the weight of the interval \(I\) appears on the right.)

The parallel corona delays the application of Lemma 4.3, this feature combined with a special function theory specific to Haar expansions for non-doubling measures, were the critical ingredients.

The parallel corona has been used to give short transparent proofs of two weight inequalities for singular integrals. See the last page of Hytönen’s survey [14] and the article of Tanaka [55].

4.1.6. It is natural to wonder if there are any \(L^p\) analogs of the main Theorem. However, there are obstructions: There is no clear version of the quasi-orthogonality argument that works in the two weight setting.

5. THE STOPPING FORM

Given an interval \(I_0\), the stopping form is

\[B_{\text{stop}}^{I_0}(f, g) := \sum_{1: I \subset I_0} \sum_{J \supset I} B_{\text{energy}}^J f \cdot \langle H_\sigma (I_0 - I_J), \Delta^w J g \rangle_w.\]

We prove this for the stopping form, which completes the proof of the inequality (4.17), and so in view of Lemma 4.16, completes the proof of the main theorem of this paper. Note that the hypotheses on \(f\) and \(g\) are that they are adapted to energy stopping intervals. (Bounded averages on \(f\) are no longer required.)

Lemma 5.2. Fix an interval \(I_0\), and let \(f\) and \(g\) be adapted to \(F_{\text{energy}}(I_0)\). Then,

\[|B_{\text{stop}}^{I_0}(f, g)| \leq \mathcal{H}||f||_w||g||_w.\]
The stopping form arises naturally in any proof of a TI theorem using Haar or other bases. In the non-homogeneous case, or in the Tb setting, where (adapted) Haar functions are important tools, it frequently appears in more or less this form. Regardless of how it arises, the stopping form is treated as an error, in that it is bounded by some simple geometric series, obtaining decay as e.g. the ratio $|J|/|I|$ is held fixed. (See for instance [35, (7.16)].)

These sorts of arguments, however, implicitly require some additional hypotheses, such as the weights being mutually $\mathcal{A}_\infty$. Of course, the two weights above can be mutually singular. There is no a priori control of the stopping form in terms of simple parameters like $|J|/|I|$, even supplemented by additional pigeonholing of various parameters.

Our method is inspired by proofs of Carleson’s Theorem on Fourier series [5, 10, 24], and has one particular precedent in the current setting, a much simpler bound for the stopping form in $T_1$ setting, where (adapted) Haar functions are important tools, it is often used in the treatment of Haar or other bases. In the $T_1$ theorem, however, implicit require some additional hypotheses, such as the weights being mutually $\mathcal{A}_\infty$. Of course, the two weights above can be mutually singular. There is no a priori control of the stopping form in terms of simple parameters like $|J|/|I|$, even supplemented by additional pigeonholing of various parameters.

Our method is inspired by proofs of Carleson’s Theorem on Fourier series [5, 10, 24], and has one particular precedent in the current setting, a much simpler bound for the stopping form in $T_1$ setting, where (adapted) Haar functions are important tools, it is often used in the treatment of Haar or other bases. In the $T_1$ theorem, however, implicit require some additional hypotheses, such as the weights being mutually $\mathcal{A}_\infty$. Of course, the two weights above can be mutually singular. There is no a priori control of the stopping form in terms of simple parameters like $|J|/|I|$, even supplemented by additional pigeonholing of various parameters.

5.1. Admissible Pairs. A range of decompositions of the stopping form necessitate a somewhat heavy notation that we introduce here. The individual summands in the stopping form involve four distinct intervals, namely $I_0$, $I$, $J_1$, and $J$. The interval $I_0$ will not change in this argument, and the pair $(I, J)$ determine $I_1$. Subsequent decompositions are easiest to phrase as actions on collections $Q$ of pairs of intervals $Q = (Q_1, Q_2)$ with $Q_1 \supseteq Q_2$. (The letter P is already taken for the Poisson integral.) And we consider the bilinear forms

$$B_Q(f, g) := \sum_{Q \in Q} E_{(Q_1)Q_2} A_{Q_1} f \cdot (H_\sigma(I_0 - (Q_1)Q_2), A_{Q_2} g)_w.$$ 

We will have the standing assumption that for all collections $Q$ that we consider are admissible.

**Definition 5.3.** A collection of pairs $Q$ is admissible if it meets these criteria. For any $Q = (Q_1, Q_2) \in Q$,

1. $Q_2 \subseteq Q_1 \subset I_0$, and both $Q_1$ and $Q_2$ are good.
2. (convexity in $Q_1$) If $Q'' \in Q$ with $Q'' = Q_2$ and $Q'' \subset I \subset Q_1$, with $I$ good, then there is a $Q' \in Q$ with $Q' = I$ and $Q_2 = Q_2$.

The first property is self-explanatory. The second property is convexity in $Q_1$, subject to goodness, holding $Q_2$ fixed, which is used in the estimates on the stopping form which conclude the argument. A third property is described below.

We exclusively use the notation $Q_k$, $k = 1, 2$ for the collection of intervals $\bigcup Q_k : Q \in Q$, not counting multiplicity. Similarly, set $Q_1 := \{(Q_1)Q_2 : Q \in Q\}$, and $\hat{Q}_1 := (Q_1)Q_2$.

3. No interval $K \subseteq Q_2$ is contained in an interval $S \in \mathcal{F}_{\mathrm{energy}}(I_0)$. (And so, no interval in $\hat{Q}_1$ is contained in an $S \in \mathcal{F}_{\mathrm{energy}}(I_0)$.)

The last requirement comes from the assumption that the functions $f$ and $g$ be adapted to $\mathcal{F}_{\mathrm{energy}}(I_0)$. We will be appealing to different Hilbertian arguments below, so we prefer to make this an assumption about the pairs rather than the functions $f, g$. The Hilbert space will be the space of good functions in $L^2(\sigma)$ and $L^2(w)$.

Typically, one only ever needs goodness of the small interval, in this case $Q_2$. We will use the term size($Q$) below, in which it will be apparent that goodness of the intervals $Q_1$ will be helpful. Namely, at this point goodness is used to as in the monotonicity principle, to estimate off-diagonal inner products involving the Hilbert transform by Poisson averages, and to regularize Poisson averages. Both are made more explicit in §5.4.

The stopping form is obtained with the admissible collection of pairs given by

$$Q_0 = \{(I, J) : J \in I, I \text{ and } J \text{ are good}, J \not\subset \bigcup \{S : S\} \}.$$
In this definition $S$ is the collection of subintervals of $I_0$ which $f$ is uniform with respect to. There holds $B_{I_0}^{\text{top}}(f, g) = B_{Q_0}(f, g)$ for $f, g$ adapted to $F_{\text{energy}}(I_0)$.

There is a very important notion of the size of $Q$.

\[
\text{size}(Q)^2 := \sup_{K \in Q_1 \cup Q_2} \frac{P(\sigma(I_0 - K), K)^2}{\sigma(K)|K|^2} \sum_{J \in Q_2 : J \subset K} \langle x, h_J^w \rangle^2_w.
\]

For admissible $Q$, there holds size($Q$) \leq H, as follows the property (3) in Definition 5.3, and Definition 4.5.

More definitions follow. Set the norm $B_Q$ of the bilinear form $Q$ to be the best constant in the inequality

\[
|B_Q(f, g)| \leq B_Q \|f\|_\sigma \|g\|_w.
\]

Thus, our goal is show that $B_Q \leq \text{size}(Q)$ for admissible $Q$, but we will only be able to do this directly in the case that the pairs $(Q_1, Q_2)$ are weakly decoupled in a collection $Q$. The relevant decoupling is precisely described in §5.4.

Say that collections of pairs $Q^j$, for $j \in \mathbb{N}$, are mutually orthogonal if on the one hand, the collections $(Q^j)_2$, of second coordinates of the pairs, are pairwise disjoint, and on the other, that the collections $(Q^j)_1$ are pairwise disjoint. The concept has to be different in the first and second coordinates of the pairs, due to the different role of the intervals $\tilde{Q}_1$ and $Q_2$, which comes up again in the next paragraph.

The meaning of mutual orthogonality is best expressed through the norm of the associated bilinear forms. Under the assumption that $B_Q = \sum_{j \in \mathbb{N}} B_{Q^j}$, and that the $(Q^j : j \in \mathbb{N})$ are mutually orthogonal, the following essential inequality holds.

\[
B_Q \leq \sqrt{2} \sup_{j \in \mathbb{N}} B_{Q^j}.
\]

Indeed, for $j \in \mathbb{N}$, let $\Pi^w_j$ be the projection onto the linear span of the Haar functions $\{h_J^w : J \in Q^j_2\}$, and use a similar notation for $\Pi^\sigma_j$. We then have the two inequalities

\[
\sum_{j \in \mathbb{N}} \|\Pi^w_j g\|_w^2 \leq \|g\|_w^2, \quad \sum_{j \in \mathbb{N}} \|\Pi^\sigma_j f\|_\sigma^2 \leq 2\|f\|_\sigma^2.
\]

Since a given interval $I$ can be in two collections $Q^j_1$, we have the factor of 2 in the second inequality. Therefore, we have

\[
|B_Q(f, g)| \leq \sum_{j \in \mathbb{N}} |B_{Q^j}(f, g)|
= \sum_{j \in \mathbb{N}} |B_{Q^j}(\Pi^\sigma_j f, \Pi^w_j g)|
\leq \sum_{j \in \mathbb{N}} B_{Q^j} \|\Pi^\sigma_j f\|_\sigma \|\Pi^w_j g\|_w \leq \sqrt{2} \sup_{j \in \mathbb{N}} B_{Q^j} \cdot \|f\|_\sigma \|g\|_w.
\]

This proves (5.5).
5.2. **The Recursive Argument.** This is the essence of the matter.

**Lemma 5.6.** [Size Lemma] An admissible collection of pairs $Q$ can be partitioned into collections $Q_{\text{large}}$ and admissible $Q_{\text{small}}$, for $t \in \mathbb{N}$ such that

$$B_Q \leq C \text{size}(Q) + (1 + \sqrt{2}) \sup_t B_{Q_{\text{small}}},$$

and

$$\sup_{t \in \mathbb{N}} \text{size}(Q_{\text{small}}) \leq \frac{1}{4} \text{size}(Q).$$

Here, $C > 0$ is an absolute constant.

The point of the lemma is that all of the constituent parts are better in some way, and that the right hand side of (5.7) involves a favorable supremum. We can quickly prove the main result of this section.

**Proof of Lemma 5.2.** The stopping form of this Lemma is of the form $B_Q(f, g)$ for admissible choice of $Q$, with $\text{size}(Q) \leq C \mathcal{H}$, as we have noted in (5.4). Define

$$\zeta(\lambda) := \sup\{B_Q : \text{size}(Q) \leq C \lambda \mathcal{H}\}, \quad 0 < \lambda \leq 1,$$

where $C > 0$ is a sufficiently large, but absolute constant, and the supremum is over admissible choices of $Q$. We are free to assume that $Q_1$ and $Q_2$ are further constrained to be in some fixed, but large, collection of intervals $I$. Then, it is clear that $\zeta(\lambda)$ is finite, for all $0 < \lambda \leq 1$. Because of the way the constant $\mathcal{H}$ enters into the definition, it remains to show that $\zeta(1)$ admits an absolute upper bound, independent of how $I$ is chosen.

It is the consequence of Lemma 5.6 that there holds

$$\zeta(\lambda) \leq C \lambda + (1 + \sqrt{2}) \zeta(\lambda/4), \quad 0 < \lambda \leq 1.$$ 

Iterating this inequality beginning at $\lambda = 1$ gives us

$$\zeta(1) \leq C + (1 + \sqrt{2}) \zeta(1/4) \leq \ldots \leq C \sum_{t=0}^{\infty} \left(\frac{1+\sqrt{2}}{4}\right)^t \leq 4C.$$

So we have established an absolute upper bound on $\zeta(1)$. □

5.3. **Proof of Lemma 5.6.** We restate the conclusion of Lemma 5.6 to more closely follow the line of argument to follow. The collection $Q$ can be partitioned into two collections $Q_{\text{large}}$ and $Q_{\text{small}}$ such that

1. $B_{Q_{\text{large}}} \leq \tau$, where $\tau := \text{size}(Q)$.
2. $Q_{\text{small}} = Q_{\text{small}}^1 \cup Q_{\text{small}}^2$.
3. The collection $Q_{\text{small}}^1$ is admissible, and $\text{size}(Q_{\text{small}}^1) \leq \frac{\tau}{4}$.
4. For a collection of dyadic intervals $\mathcal{L}$, the collection $Q_{\text{small}}^2$ is the union of mutually orthogonal admissible collections $Q_{\text{small}}^{2, L}$, for $L \in \mathcal{L}$, with

$$\text{size}(Q_{\text{small}}^{2, L}) \leq \frac{\tau}{4}, \quad L \in \mathcal{L}.$$ 

Thus, we have by inequality (5.5) for mutually orthogonal collections,

$$B_Q \leq B_{Q_{\text{large}}} + B_{Q_{\text{small}}^1 \cup Q_{\text{small}}^2}$$

$$\leq B_{Q_{\text{large}}} + B_{Q_{\text{small}}^1} + B_{Q_{\text{small}}^2}$$

$$\leq C \tau + (1 + \sqrt{2}) \max\{B_{Q_{\text{small}}^1}, \sup_{L \in \mathcal{L}} B_{Q_{\text{small}}^{2, L}}\}.$$
This, with the properties of size listed above prove Lemma 5.6 as stated, after a trivial re-indexing.

In a manner similar to the proof of the functional energy inequality, there is an induced measure on the upper half-plane that is relevant to our considerations. This time it is given by
\[ \mu_Q = \mu := \sum_{J \in Q_2} \langle x, h_J^w \rangle^2_w \delta_{[x_J, |J|]}, \quad x_J \text{ is the center of } J. \]

The tent over $L$ is the triangular region $T_L := \{(x, y) : |x - x_L| \leq |L| - y\}$, so that
\[ \mu(T_L) = \sum_{J \in Q_2} \langle x, h_J^w \rangle^2_w. \]

Observe that
\[ \text{size}(Q)^2 = \sup_{K \in Q_1 \cup Q_2} \frac{P(\sigma(I_0 - K), K)^2}{\sigma(K) |K|^2} \mu(T_K). \]

All else flows from this construction of a subset $L$ of dyadic subintervals of $I_0$. The initial intervals in $L$ are the minimal intervals $L \in \tilde{Q}_1 \cup Q_2$ such that
\[ (5.8) \quad \frac{P(\sigma(I_0 - L), L)^2}{|L|^2} \mu(T_L) \geq \frac{\tau^2}{16} \sigma(L). \]

Since $\text{size}(Q) = \tau$, there are such intervals $L$.

Initialize $S$ (for 'stock' or 'supply') to be all the dyadic intervals in $\tilde{Q}_1 \cup Q_2$ which strictly contain some interval in $L$. In the recursive step, let $L'$ be the minimal elements $S \in S$ such that
\[ (5.9) \quad \mu(T_S) \geq \rho \sum_{L \in S: L \subset S \text{ is maximal}} \mu(T_L), \quad \rho = \frac{17}{16}. \]

(The inequality would be trivial if $\rho = 1$.) If $L'$ is empty the recursion stops. Otherwise, update $L \leftarrow L \cup L'$, and $S \leftarrow \{K \in S : K \not\subset L \forall L \in L\}$. See Figure 8.

Once the recursion stops, report the collection $L$. It has this crucial property: For $L \in L$, and integers $t \geq 1$,
\[ (5.10) \quad \sum_{L' : \pi^L_{L'} \subset L} \mu(T_{L'}) \leq \rho^{-t} \mu(T_L). \]

Indeed, in the case of $t = 1$, is a criteria for membership in $L$, and a simple induction proves the statement for all $t \geq 1$.

The decomposition of $Q$ is based upon the relation of the pairs to the collection $L$, namely a pair $\tilde{Q}_1, Q_2$ can (a) both have the same parent in $L$; (b) have distinct parents in $L$; (c) $Q_2$ can have a parent in $L$, but not $\tilde{Q}_1$; and (d) $Q_2$ does not have a parent in $L$.

A particularly vexing aspect of the stopping form is the linkage between the martingale difference on $g$, which is given by $J$, and the argument of the Hilbert transform, $I_0 - I_J$. The ‘large’ collections constructed below will, in a certain way, decouple the $J$ and the $I_0 - I_J$, enough so that norm of the associated bilinear form can be estimated by the size of $Q$.

In the ‘small’ collections, there is however no decoupling, but critically, the size of the collections is smaller, and we only have to estimate the maximal operator norm among the small collections.
Pairs comparable to $L$. Define
\[
Q_{L,t} := \{ Q \in Q : \pi_L \tilde{Q}_1 = \pi_t Q_2 = L \}, \quad L \in \mathcal{L}, \ t \in \mathbb{N}.
\]
These are admissible collections, as the convexity property in $Q_1$, holding $Q_2$ constant, is clearly inherited from $Q$. Now, observe that for each $t \in \mathbb{N}$, the collections $\{Q_{L,t} : L \in \mathcal{L}\}$ are mutually orthogonal. The collection of intervals $(Q_{L,t})_L$ are obviously disjoint in $L \in \mathcal{L}$, with $t \in \mathbb{N}$ held fixed. And, since membership in these collections is determined in the first coordinate by the interval $\tilde{Q}_1$, and the two children of $Q_1$ can have two different parents in $\mathcal{L}$, a given interval $I$ can appear in at most two collections $(Q_{L,t})_L$, as $L \in \mathcal{L}$ varies, and $t \in \mathbb{N}$ held fixed.

Define $Q_{L,1}^{\text{small}}$ to be the union over $L \in \mathcal{L}$ of the collections
\[
Q_{L,1}^{\text{small}} := \{ Q \in Q_{L,1} : \tilde{Q}_1 \neq L \}.
\]
Note in particular that we have only allowed $t = 1$ above, and $\tilde{Q}_1 = L$ is not allowed. For these collections, we need only verify that

\begin{equation}
\text{size}(Q_{L,1}^{\text{small}}) \leq \sqrt{(\rho - 1) \cdot \tau} = \frac{\tau}{4}, \quad L \in \mathcal{L}, \ t \in \mathbb{N}.
\end{equation}

\textbf{Lemma 5.11.} There holds

\begin{equation}
(5.12) \quad \text{size}(Q_{L,1}^{\text{small}}) \leq \sqrt{(\rho - 1) \cdot \tau} = \frac{\tau}{4}, \quad L \in \mathcal{L}, \ t \in \mathbb{N}.
\end{equation}

\textbf{Proof.} An interval $K \in (Q_{L,1}^{\text{small}}) \cup Q_2$ is not in $\mathcal{L}$, by construction. Suppose that $K$ does not contain any interval in $\mathcal{L}$. By the selection of the initial intervals in $\mathcal{L}$, the minimal intervals in $\tilde{Q}_1 \cup Q_2$ which satisfy (5.8), it follows that the interval $K$ must fail (5.8). And so we are done.

Thus, $K$ contains some element of $\mathcal{L}$, whence the inequality (5.9) must fail. Namely, rearranging that inequality, and using the measure $\mu$ associated with $Q_{L,1}^{\text{small}}$,

\[
\mu_{Q_{L,1}^{\text{small}}} (T_L) \leq (\rho - 1) \sum_{L' \in \mathcal{L} : L' \subset K, L' \text{ is maximal}} \mu(T_L)
\]

\[
\leq \frac{1}{16} \mu(T_L) \leq \frac{\tau^2}{16} \cdot \frac{|K|^2 \cdot \sigma(K)}{P(\sigma(L-K), K)^2}.
\]

Here, note that we begin with the measure $\mu_{Q_{L,1}^{\text{small}}}$; use $\rho = 1 + \frac{1}{16}$; and the last inequality follows from the definition of size. This finishes the proof of (5.12). \qed
The collections below are the first contribution to $Q^{\text{large}}$. Take $Q^{\text{large}}_1 := \bigcup(Q^{\text{large}}_{L,t} : L \in \mathcal{L})$, where $Q^{\text{large}}_{L,1} := \{Q \in Q_{L,1} : \bar{Q}_1 = L\}$.

Note that Lemma 5.19 applies to this Lemma, take the collection $S$ of that Lemma to be $\{L\}$, and the quantity $\eta$ in (5.20) satisfies $\eta \leq \tau = \text{size}(Q)$, by (5.21). From the mutual orthogonality (5.5), we then have

$$B_{Q^{\text{large}}_1} \leq \sqrt{2} \sup_{L \in \mathcal{L}} B_{Q^{\text{large}}_{L,1}} \lesssim \tau.$$  

The collections $Q^{\text{large}}_{L,t},$ for $L \in \mathcal{L}$, and $t \geq 2$ are the second contribution to $Q^{\text{large}}$, namely

$$Q^{\text{large}}_2 := \bigcup_{L \in \mathcal{L}} \bigcup_{t \geq 2} Q_{L,t}.$$  

For them, we need to estimate $B_{Q^{\text{large}}_2}$.

**Lemma 5.13.** There holds $B_{Q^{\text{large}}_2} \lesssim \rho^{-1/2}\tau$.

From this, we can conclude from (5.5) that

$$B_{Q^{\text{large}}_2} \leq \sum_{t \geq 2} B_{\bigcup(Q_{L,t} : L \in \mathcal{L})} \leq \sqrt{2} \sum_{t \geq 2} \sup_{L \in \mathcal{L}} B_{Q_{L,t}} \lesssim \tau \sum_{t \geq 2} \rho^{-1/2} \lesssim \tau.$$  

**Proof.** For $L \in \mathcal{L}$, let $S_L$, the $\mathcal{L}$-children of $L$. For each $Q \in Q_{L,t}$, we must have $Q_2 \subset \pi_{S_L} Q_2 \subset \bar{Q}_1$. Then, divide the collection $Q_{L,t}$ into three collections $Q_{L,t}^\ell$, $\ell = 1,2,3$, where

$$Q_{L,t}^1 := \{Q \in Q_{L,t} : Q_2 \subset \pi_{S_L} Q_2\},$$  

$$Q_{L,t}^2 := \{Q \in Q_{L,t} : Q_2 \not\subset \pi_{S_L} Q_2 \} = \{Q \in Q_{L,t} : Q_2 \not\subset \bar{Q}_1\},$$  

and $Q_{L,t}^3 := Q_{L,t} \setminus (Q_{L,t}^1 \cup Q_{L,t}^2)$ is the complementary collection. Notice that $Q_{L,t}^1$ equals the whole collection $Q_{L,t}$ for $t > r + 1$.

We treat them in turn. The collections $Q_{L,t}^1$ fit the hypotheses of Lemma 5.19, just take the collection of intervals $S$ of that Lemma to be $S_L$. It follows that $B_{Q_{L,t}^1} \leq \beta(t)$, where the latter is the best constant in the inequality

$$\sum_{J \in \mathcal{Q}_{L,1}} \sum_{J \in \mathcal{K}} P(\sigma(I_0 - K), J)^2 \frac{X}{||J||} \frac{h^w}{w} \leq \beta(t)^2 \sigma(K), \quad K \in S_L, \ L \in \mathcal{L}, \ t \geq 2.$$  

We will prove the estimate below, which is clearly summable in $t \in \mathbb{N}$ to the estimate we want.

**Lemma 5.15.** There holds $\beta(t) \leq \rho^{-1/2}\tau$.

**Proof.** We have the estimate without decay in $t$, $\beta(t) \leq \text{size}(Q)$, as follows from (5.21). Use this estimate for $1 \leq t \leq r + 3$, say. In the case of $t > r + 3$, the essential property is (5.10). The left hand side of (5.14) is dominated by the sum below. Note that we index the sum first over $L'$, which are $r + 1$-fold $\mathcal{L}$-children of $K$, whence $L' \in K$, followed by $t - r - 2$-fold $\mathcal{L}$-children of $L'$.  

$$\sum_{L' \in \mathcal{L}} \sum_{\pi_{L'} L'' = K} \sum_{K'' \in Q_{L'} : L'' \subset L''} P(\sigma(I_0 - K), J)^2 \frac{X}{||J||} \frac{h^w}{w}.$$  

We will prove the estimate below, which is clearly summable in $t \in \mathbb{N}$ to the estimate we want.
\[ \sum_{L \in \mathcal{L}, \pi_{L_{i}^{+}}^{+} L' = K} \frac{P(|I_{0} - K)|, L')^2}{|L'|^2} \sum_{L'' \in \mathcal{L}, \pi_{L_{i}^{+}}^{+} L'' = L'} \mu(T_{L''}) \]

\[ \sum_{L \in \mathcal{L}, \pi_{L_{i}^{+}}^{+} L' = K} \frac{P(|I_{0} - K)|, L')^2}{|L'|^2} \mu(T_{L'}) \]

\[ \leq \rho^{-t+\tau+2} \sum_{L' \in \mathcal{L}, \pi_{L_{i}^{+}}^{+} L' = K} \frac{P(|I_{0} - K)|, L')^2}{|L'|^2} \mu(T_{L'}) \]

\[ \leq \rho^{-t}\tau^2 \sum_{L' \in \mathcal{L}, \pi_{L_{i}^{+}}^{+} L' = K} \sigma(L') \leq \tau^2 \rho^{-t} \sigma(K). \]

We have also used (5.18), and then the central property (5.10) following from the construction of \( \mathcal{L} \), finally appealing to the definition of size. Hence, \( \beta(t) \leq \tau \rho^{-t/2} \). This completes the analysis of \( Q_{L,t}^1 \).

□

We need only consider the collections \( Q_{L,t}^2 \) for \( 1 \leq t \leq r + 1 \), and they fall under the scope of Lemma 5.25. A variant of (5.21) shows that \( B_{Q_{L,t}^2} \leq \tau \). Similarly, we need only consider the collections \( Q_{L,t}^3 \) for \( 1 \leq t \leq r + 1 \). It follows that we must have \( 2^r \leq |Q_1|/|Q_2| \leq 2^{2r+2} \). Namely, this ratio can take only one of a finite number of values, implying that Lemma 5.27 applies easily to this case to complete the proof.

□

**Pairs not strictly comparable to \( \mathcal{L} \).** It remains to consider the pairs \( Q \in \mathcal{Q} \) such that \( \tilde{Q}_1 \) does not have a parent in \( \mathcal{L} \). The collection \( Q_{2,\text{small}}^\text{small} \) is taken to be the (much smaller) collection

\[ Q_{2,\text{small}}^\text{small} := \{ Q \in \mathcal{Q} : \text{Q}_2 \text{ does not have a parent in } \mathcal{L} \}. \]

Observe that size\( (Q_{2,\text{small}}^\text{small}) \leq \sqrt{(\rho - 1)} \tau \leq \frac{\tau}{4} \). This is as required for this collection.\(^5\)

**Proof.** Suppose \( \eta < \text{size}(Q_{2,\text{small}}^\text{small}) \). Then, there is an interval \( K \in (Q_{1,\text{small}}^\text{small})_1 \cup (Q_{2,\text{small}}^\text{small})_2 \) so that

\[ \eta^2 \sigma(K) \leq \frac{P(|I_{0} - K)|, K)^2}{|K|^2} \mu_{Q_{2,\text{small}}^\text{small}}(T_K). \]

Suppose that \( K \) does not contain any interval in \( \mathcal{L} \). It follows from the initial intervals added to \( \mathcal{L} \), see (5.8), that we must have \( \eta \leq \frac{\tau}{4} \).

Thus, \( K \) contains an interval in \( \mathcal{L} \). This means that \( K \) must fail the inequality (5.9). Therefore, we have

\[ \eta^2 \sigma(K) \leq (\rho - 1) \frac{P(|I_{0} - K)|, K)^2}{|K|^2} \mu(T_K) \leq \frac{\tau^2}{16} \sigma(K). \]

This relies upon the definition of size, and proves our claim.

□

For the pairs not yet in one of our collections, it must be that \( Q_2 \) has a parent in \( \mathcal{L} \), but not \( \tilde{Q}_1 \). Using \( \mathcal{L}^* \), the maximal intervals in \( \mathcal{L} \), divide them into the three collections

\[ Q_{3,\text{large}}^\text{large} := \{ Q \in \mathcal{Q} : Q_2 \notin \pi_{\mathcal{L}^*} Q_2 \subset \tilde{Q}_1 \}, \]

\[ Q_{4,\text{large}}^\text{large} := \{ Q \in \mathcal{Q} : Q_2 \notin \pi_{\mathcal{L}^*}, Q_2 \notin \tilde{Q}_1 \}. \]

\(^5\)The collections \( Q_{1,\text{small}}^\text{small} \) and \( Q_{2,\text{small}}^\text{small} \) are also mutually orthogonal, but this fact is not needed for our proof.
\[ Q_3^{\text{large}} := \{ Q \in \mathcal{Q} : Q_2 \notin \mathcal{P}_1, Q_2 \subseteq \tilde{Q}_1, \text{ and } \pi_L \cdot Q_2 \notin \tilde{Q}_1 \}. \]

Observe that Lemma 5.19, with (5.21), gives

\[ (5.16) \quad B_{Q_3^{\text{large}}} \leq \tau. \]

Take the collection \( \mathcal{S} \) of Lemma 5.19 to be \( L^* \).

Observe that Lemma 5.25 applies to show that the estimate (5.16) holds for \( Q_4^{\text{large}} \). Take \( \mathcal{S} \) of that Lemma to be \( L^* \). The estimate from Lemma 5.25 is given in terms of \( \eta \), as defined in (5.26). But, is at most \( \tau \).

In the last collection, \( Q_5^{\text{large}} \), notice that the conditions placed upon the pair implies that \( |Q_1| \leq 2^{2r+2} |Q_2| \), for all \( Q \in Q_5^{\text{large}} \). It therefore follows from a straightforward application of Lemma 5.27, that (5.16) holds for this collection as well.

5.4. Upper Bounds on the Stopping Form. We prove upper bounds on the norm of the stopping form in a situation in which there is some decoupling between the martingale difference on \( g \), and the argument of the Hilbert transform. First, an elementary observation.

**Proposition 5.17.** For intervals \( J \subset L \Subset K \), with \( L \) either good, or the child of a good interval,

\[ (5.18) \quad \frac{\mathbb{P}(\sigma(I_0 - K), J)}{|J|} \simeq \frac{\mathbb{P}(\sigma(I_0 - K), L)}{|L|}. \]

**Proof.** The property of interval \( I \) being good, says that if \( I \subset \bar{I} \), and \( 2^{r-1} |I| \leq |\bar{I}| \), then the distance of either child of \( I \) to the boundary of \( \bar{I} \) is at least \( |I|^{\epsilon} |\bar{I}|^{1-\epsilon} \). Thus, in the case that \( L \) is the child of a good interval, the parent \( \bar{\tilde{I}} \) of \( L \) is contained in \( K \), and \( 2^{r-1} |\bar{\tilde{I}}| \leq |K| \), so by the definition of goodness,

\[ \text{dist}(J, I_0 - K) \geq \text{dist}(L, I_0 - K) \]

\[ \geq |L|^{\epsilon} |K|^{1-\epsilon} \geq 2^{(1-\epsilon)} |L|. \]

The same inequality holds if \( L \) is good. Then, one has the equivalence above, by inspection of the Poisson integrals. \( \square \)

**Lemma 5.19.** Let \( \mathcal{S} \) be a collection of pairwise disjoint intervals in \( I_0 \). Let \( \mathcal{Q} \) be admissible such that for each \( Q \in \mathcal{Q} \), there is an \( S \in \mathcal{S} \) with \( Q_2 \Subset S \subset Q_1 \). Then, there holds

\[ |B_{\mathcal{Q}}(f, g)| \leq \eta \|f\|_w \|g\|_w, \]

(5.20) where \( \eta^2 := \sup_{S \in \mathcal{S}} \frac{1}{\sigma(S)} \sum_{J \in Q_2 : J \Subset S} \mathbb{P}(\sigma(I_0 - S), J)^2 \langle x, h^w \rangle_w^2. \)

It is useful to note that \( \eta \) is always smaller than the size: For \( S \in \mathcal{S} \), let \( J^* \) be the maximal intervals \( J \in Q_2 \) with \( J \Subset S \), and note that (5.18) applies to see that

\[ \sum_{J \in Q_2 : J \Subset S} \mathbb{P}(\sigma(I_0 - S), J)^2 \langle x, h^w \rangle_w^2 = \sum_{J^* \in J^*} \sum_{J \in Q : J \subseteq J^*} \mathbb{P}(\sigma(I_0 - S), J)^2 \langle x, h^w \rangle_w^2 \]

\[ \leq \sum_{J^* \in J^*} \frac{\mathbb{P}(\sigma(I_0 - S), J^*)^2}{|J^*|^2} \sum_{J \in Q : J \subseteq J^*} \langle x, h^w \rangle_w^2 \]

(5.21) \[ \leq \sum_{J^* \in J^*} \sigma(J^*) \leq \text{size}(Q) \sigma(S). \]
Proof. An interesting part of the proof is that it depends very much on cancellative properties of the martingale differences of $f$. (Absolute values must be taken outside the sum defining the stopping form!) This argument will invoke the stopping data, and part of the Hilbert-Poisson exchange argument.

Assume, as we can, that the Haar support of $f$ is contained in $Q_1$. Take $\mathcal{F}$ and $\alpha_1(\cdot)$ to be stopping data defined in this way: First, add to $\mathcal{F}$ the interval $I_0$, and set $\alpha_1(I_0) := \mathbb{E}_0^f|f|$. Inductively, if $F \in \mathcal{F}$ is minimal, add to $\mathcal{F}$ the maximal children $F'$ such that $\alpha_1(F') := \mathbb{E}_0^g|f| > 4\alpha_1(F)$. This is a simple form of the stopping data construction in §4. In particular quasi-orthogonality (4.9) holds.

Write the bilinear form as

$$ B_Q(f, g) = \sum J \langle H_\sigma \varphi_1, \Delta_j^w g \rangle_w $$

(5.22) where $\varphi_1 := \sum_{Q \in \mathcal{Q} : Q_2 = J} \mathbb{B}_0^Q \Delta_0^Q \cdot (I_0 - \tilde{Q}_1)$. The function $\varphi_1$ is well-behaved, as we now explain. At each point $x$ with $\varphi_1(x) \neq 0$, the sum above is over pairs $Q$ such that $Q_2 = J$ and $x \in I_0 - \tilde{Q}_1$. By the convexity property of admissible collections, the sum is over consecutive (good) martingale differences of $f$. The basic telescoping property of these differences shows that the sum is bounded by the stopping value $\alpha_1(\tau_F J)$. Let $I^*$ be the maximal interval of the form $Q_1$ with $x \in I_0 - \tilde{Q}_1$, and let $I_*$ be the child of the minimal such interval which contains $J$. Then,

$$ |\varphi_1(x)| = \left| \sum_{Q \in \mathcal{Q} : Q_2 = J} \mathbb{B}_0^Q \Delta_0^Q f(x) \right| $$

(5.23) where $|S - \mathbb{E}_0^J f| \leq \alpha_1(\tau_F J)(I_0 - S)$, where $S$ is the $S$-parent of $J$.

We can estimate as below, for $F \in \mathcal{F}$:

$$ \Xi(F) := \left| \sum_{Q \in \mathcal{Q} : \pi_S Q_2 = F} \mathbb{B}_Q \Delta_0^Q f \cdot \langle H_\sigma (I_0 - \tilde{Q}_1), \Delta_j^w g \rangle_w \right| $$

(5.22) $\leq \alpha_1(F) \left[ \sum_{S \in \mathcal{S}} \sum_{J \in \mathcal{Q}_2} P(\sigma(I_0 - S), J) \left| \frac{x}{|J|}, \Delta_j^w g \right\rangle_w \right|$

$\leq \alpha_1(F) \left[ \sum_{S \in \mathcal{S}} \sum_{J \in \mathcal{Q}_2} P(\sigma(I_0 - S), J) \frac{x^2}{|J|^2} \frac{h_j^w}{w} \times \sum_{J \in \mathcal{Q}_2} \hat{g}(J)^2 \right]^{1/2}$

(5.20) $\leq \text{size}(\mathcal{Q}) \alpha_1(F) \left[ \sum_{S \in \mathcal{S}} \sigma(S) \times \sum_{J \in \mathcal{Q}_2} \hat{g}(J)^2 \right]^{1/2}$

$\leq \text{size}(\mathcal{Q}) \alpha_1(F) \sigma(F)^{1/2} \left[ \sum_{J \in \mathcal{Q}_2 : \pi_F J = F} \hat{g}(J)^2 \right]^{1/2}$. 

The top line follows from (5.22). In the second, we appeal to (5.23) and monotonicity principle, the latter being available to us since \( J \subset S \) implies \( J \subset S \), by hypothesis. We also take advantage of the strong assumptions on the intervals in \( Q_2 \): If \( J \in Q_2 \), we must have \( \alpha_{JF} = \alpha_{JF}(\pi_S) \). The third line is Cauchy–Schwarz, followed by the appeal to the hypothesis (5.20), while the last line uses the fact that the intervals in \( S \) are pairwise disjoint.

The quasi-orthogonality argument (4.9) completes the proof, namely we have

\[(5.24) \quad \sum_{F \in \mathcal{F}} \Xi(F) \leq \text{size}(\mathcal{Q}) ||f||_\sigma ||g||_w.
\]

\[\square\]

**Lemma 5.25.** Let \( S \) be a collection of pairwise disjoint intervals in \( I_0 \). Let \( \mathcal{Q} \) be admissible such that for each \( Q \in \mathcal{Q} \), there is an \( S \in \mathcal{S} \) with \( Q_2 \subset S \in \mathcal{Q}_1 \). Then, there holds

\[|B_\mathcal{Q}(f, g)| \leq \eta \|f\|_\sigma \|g\|_w,
\]

(5.26) where \( \eta^2 := \sup_{S \in \mathcal{S}} \frac{P(\sigma(Q_1 - \pi_Q S), S)^2}{\sigma(S)|S|^2} \sum_{J \in Q_2 : J \subset S} \langle x, h_J^w \rangle_w^2.\]

**Proof.** Construct stopping data \( \mathcal{F} \) and \( \alpha_{f}(\cdot) \) as in the proof of Lemma 5.19. The fundamental inequality (5.23) is again used. Then, by the monotonicity principle (3.15), there holds for \( F \in \mathcal{F}, \)

\[\Xi(F) := \left| \sum_{Q \in \mathcal{Q} : \pi_F Q = F} E_{Q_2} \Delta_{Q_1}^r f \cdot \langle H_{\sigma}(I_0 - \tilde{Q}_1), \Delta_{Q_2}^w g \rangle_w \right|
\]

\[\leq \alpha_{f}(F) \sum_{S \in \mathcal{S} : \pi_F S = F} P(\sigma(Q_1 - \pi_Q S), S) \sum_{J \in Q_2 : J \subset S} \langle \frac{x}{|S|}, h_J^w \rangle_w^2 \cdot |\hat{g}(J)|
\]

\[\leq \alpha_{f}(F) \left[ \sum_{S \in \mathcal{S} : \pi_F S = F} P(\sigma(Q_1 - \pi_Q S), S)^2 \sum_{J \in Q_2 : J \subset S} \langle \frac{x}{|S|}, h_J^w \rangle_w^2 \times \sum_{J \in Q_2 : J \subset S} |\hat{g}(J)|^2 \right]^{1/2}
\]

\[\leq \eta \alpha_{f}(F) \left[ \sum_{S \in \mathcal{S} : \pi_F S = F} \sigma(S) \times \sum_{J \in Q_2 : J \subset S} |\hat{g}(J)|^2 \right]^{1/2}.
\]

After the monotonicity principle (3.15), we have used Cauchy-Schwarz, and the definition of \( \eta \). The quasi-orthogonality argument (4.8) then completes the analysis of this term, see (5.24). \[\square\]

The last Lemma that we need is elementary, and is contained in the methods of [35].

**Lemma 5.27.** Let \( u \geq r + 1 \) be an integer, and \( \mathcal{Q} \) be an admissible collection of pairs such that \(|Q_1| = 2^u |Q_2| \) for all \( Q \in \mathcal{Q} \). Then holds

\[|B_\mathcal{Q}(f, g)| \leq \text{size}(\mathcal{Q}) ||f||_\sigma \|g\|_w.
\]

**Proof.** Recall the form of the stopping form in (5.1). Observe, from inspection of the definition of the Haar function (2.1), that

\[|E_{I_1} \Delta_{I_1}^r f| \leq \frac{|\hat{f}(I_1)|}{\sigma(I_1)^{1/2}}.
\]
Then, an elementary application of the monotonicity principle gives us

$$|B_\mathcal{Q}(f, g)| \leq \sum_{I \in \mathcal{Q}_1} |\hat{f}(I)| \sum_{J : (I, J) \in \mathcal{Q}} \sigma(I_J)^{-1/2} P(\sigma(0 - I_J), J) \langle \frac{x}{|I|}, h|w| \rangle \hat{g}(J) |J|$$

$$\leq \|f\|_\sigma \left[ \sum_{I \in \mathcal{Q}_1} \left( \sum_{J : (I, J) \in \mathcal{Q}} \frac{1}{\sigma(I_J)} P(\sigma(0 - I_J), J) \langle \frac{x}{|I|}, h|w| \rangle \hat{g}(J) \right)^2 \right]^{1/2}$$

$$\leq \text{size}(\mathcal{Q}) \|f\|_\sigma \|g\|_w$$

This follows immediately from Cauchy-Schwarz, and the fact that for each $J \in \mathcal{Q}_2$, there is a unique $I \in \mathcal{Q}_1$ such that the pair $(I, J)$ contribute to the sum above. 

5.5. **Context and Discussion.**

5.5.1. In the functional energy inequality, one ‘stops’ at $\sigma$-Carleson family of intervals, where as in the stopping form, every interval is ‘stopping’: the Haar function applied to $g$ and the argument of the Hilbert transform are coupled. This leads to many complications, such as the functional energy inequality has a nearly intrinsic formulation, while the stopping form does not. The proof herein succeeds because the notion of size approximates the operator norm of the stopping form. Moreover, the ‘large’ portions of the stopping form, there is a decoupling that takes place.

5.5.2. It is very interesting that one can prove unconditional results about the two weight Hilbert transform, following the techniques in [22], without solving the local problem.

6. **Elementary Estimates**

This section is devoted to the proof of Lemma 4.3. The estimates fall into many subcases, and are of a more classical nature, albeit the $A_2$ assumption is critical. (In fact, all the estimates in this section depend only on the half-Poisson $A_2$ hypothesis, but this is not systematically tracked in the notation.) In addition, all estimates should be interpreted as uniform over all smooth truncations. Some of these are off-diagonal estimates, for which the smooth truncations are important. The uniformity over truncations is however suppressed in notation.

First some basic estimates are collected. This is property of good intervals, which can be effectively used in non-critical situations.

**Lemma 6.1.** For three intervals $J, I, I' \in \mathcal{D}$ with $J \subset I \subset I'$, $|J| = 2^{-s}|I|$, with $s \geq r$ and $J$ good, then

$$(6.2) \quad P(\sigma \cdot (I' - I), J) \leq 2^{-s(1-\varepsilon)} P(\sigma \cdot I', I).$$

**Proof.** Note that for $x \in I' - I$ we have

$$\text{dist}(x, J) \geq |I|^{1-\varepsilon} |J|^\varepsilon = 2^{s(1-\varepsilon)} |J|. $$

Using this in the definition of the Poisson integral, we get

$$P(\sigma \cdot (I' - I), J) \leq 2 \int_{I' - I} \frac{|J|}{\text{dist}(x, J)^2} \sigma(dx)$$

$$\leq \frac{|J|}{{|I|}} \int_{I' - I} \frac{|I|}{(|I| + \text{dist}(x, I))^2} \sigma(dx)$$

$$\leq 2^{-s(1-2\varepsilon)} \int_{I' - I} \frac{|I|}{(|I| + \text{dist}(x, I))^2} \sigma(dx) = 2^{-s(1-2\varepsilon)} P(\sigma(I' - I), I).$$
Proposition 6.3. Suppose that two intervals $I, J \in \mathcal{D}$ satisfy $|I| \geq |J|$, and $3I \cap J = \emptyset$, then

$$ \sup_{0 < \alpha < \beta} |\langle H_{\alpha, \beta}(\sigma I), h_J^w \rangle_w| \leq \sigma(I) \sqrt{w(J)} \frac{|J|}{(||| + \text{dist}(I, J))}^2 $$

Proof. We use (3.12), thus $0 < \alpha < \beta$ are choices of truncation levels. Since $h_J^w$ has $w$-integral zero, estimate as below, where $x_J$ is the center of $J$.

$$ |\langle H_{\alpha, \beta}(\sigma I), h_J^w \rangle_w| = \left| \int_{J} \int_{J} K_{\alpha, \beta}(y - x) \cdot h_J^w(x) \, w(dx)\sigma(dy) \right| $$

$$ = \left| \int_{J} \int_{J} \left(K_{\alpha, \beta}(y - x) - K_{\alpha, \beta}(y - x_J)\right) h_J^w(x) \, w(dx)\sigma(dy) \right| $$

$$ \leq \left| \int_{J} \int_{J} \frac{|J|}{(||| + \text{dist}(I, J))} |h_J^w(x)| \, w(dx)\sigma(dy) \right|. $$

The Lemma follows by inspection.

Proposition 6.5. Suppose that two intervals $I, J \in \mathcal{D}$ satisfy $2^s|J| = |I|$, where $s > r$, the interval $J$ is good, and $J \subset 3I \setminus I$, then

$$ \sup_{0 < \alpha < \beta} |\langle H_{\alpha, \beta}(\sigma I), h_J^w \rangle_w| \leq 2^{-(1-2\varepsilon)s} \sigma(I) \sqrt{w(J)|J|}^{-1} $$

Proof. Under the assumption of the Lemma, the proof of Proposition 6.3 holds, supplying the estimate of that Lemma. But, the extra assumption that $J$ is good implies that $\text{dist}(J, I) > 2^{s(1-\varepsilon)|J|}$, and then the estimate follows by inspection.

6.1. The Weak Boundedness Inequality. The following inequality is a weak-boundedness inequality, a consequence of the $A_2$ inequality.

Proposition 6.7. There holds for all intervals $I, J$ with no point masses at their endpoints,

$$ \sup_{0 < \alpha < \beta} |\langle H_{\alpha, \beta}(\sigma f \cdot I), g \cdot J \rangle_w| \leq A_2^{1/2} \|f\|_w \|g\|_w. $$

The constant on the right can in fact be taken as follows. For a point $a$ that separates the interiors of $I$ and $J$, with $I$ to the left of $a$,

$$ \sup_{r>0} P(\sigma 1_{(-\infty, a)}, (a, a + r)) \frac{w(a, a + r)}{r} + P(\sigma 1_{[a, \infty)}, (a, a + r)) \frac{\sigma(a - r, a)}{r}. $$

In particular, for arbitrary intervals $I$ and $J$ with no point masses at the endpoints,

$$ |\langle H_\sigma I, J \rangle_w| \leq \mathcal{H}(\sigma(I)w(J))^{1/2} $$

It is useful to note that the global integrability of indicators is then a consequence of the $A_2$ and interval testing conditions.

Corollary 6.11. There holds for intervals $I$ and $J$ with no point mass at their endpoints,

$$ \|H_\sigma I\|_w \leq \mathcal{H}(\sigma(I))^{1/2}, \quad \text{and} \quad \|H_w I\|_{\sigma} \leq \mathcal{H}(w(I))^{1/2}. $$
Proof. The two sets of inequalities are of course dual. The $L^2(w)$ norm of $H\sigma I$, restricted to just the interval $I = (a, b)$ is the testing condition. Restricted to $[a, \infty)$, is the inequality (6.8), and likewise for $(-\infty, b]$. So the proof is complete.

Since the intervals are disjoint, there is no possibility of cancellation in the estimate, and it therefore is closely relate to the Hardy inequality. In the two weight setting, this has been characterized by Muckenhoupt [29].

**Theorem F.** For weights $\hat{w}$ and $\sigma$ supported on $\mathbb{R}_+$.

\[
\left\| \int_0^x f \sigma(dy) \right\|_{\hat{w}} \leq B \|f\|_{\sigma},
\]

(6.12) where $B^2 \simeq \sup_{0<r} \int_r^\infty \hat{w}(dx) \times \int_0^r \sigma(dy)$.

For the sake of completeness, we recall Muckenhoupt’s proof of this result. This preparation is proved by integration by parts.

**Proposition 6.13.** Let $\phi$ be an increasing function on $[0, \infty)$, with $\phi(0) = 0$ and $\phi$ strictly positive on $(0, \infty)$. Then,

\[
\int_0^\infty \phi(t)^{-1/2} d\phi(t) = 2\phi(x)^{1/2}.
\]

(6.14)

**Proof of Prior Result F.** We are free to assume that the function $\phi(x) = \sigma(0, x)$ is strictly positive on $(0, \infty)$. Then, multiply and divide by $\phi(x)^{1/4}$, and use Cauchy–Schwarz to see that

\[
\left\| \int_0^x f \sigma(dy) \right\|_{\hat{w}}^2 \leq \int_0^\infty \int_0^x f(y) \phi(y)^{1/2} \sigma(dy) \cdot \int_0^x \phi(y)^{-1/2} \sigma(dy) \hat{w}(dx)
\]

\[
= 2 \int_0^\infty \int_0^x f(y) \phi(y)^{1/2} \sigma(dy) \cdot \phi(x)^{1/2} \hat{w}(dx)
\]

\[
= 2 \int_0^\infty f(y) \phi(y)^{1/2} \int_0^x \phi(x)^{1/2} \hat{w}(dx) \sigma(dy)
\]

Above, we have used (6.14), and then Fubini. Concentrate on the inner integral. Using the hypothesis, it is at most

\[
B \int_y^\infty \left[ \int_x^\infty \hat{w}(dt) \right]^{-1/2} \hat{w}(dx) = 2B \left[ \int_y^\infty \hat{w}(dt) \right]^{1/2}
\]

again by (6.14). But now our assumption gives us

\[
\left\| \int_0^x f \sigma(dy) \right\|_{\hat{w}}^2 \leq 4B \int_0^\infty f(y) \phi(y)^{1/2} \int_y^\infty \hat{w}(dt)^{1/2} \sigma(dy) \leq 4B^2 \|f\|^2_{\sigma}.
\]

The proof is complete.

**Proof of Proposition 6.7.** Interval testing and (6.8) prove the estimate (6.10), so we turn to the proof of (6.8).

After a translation, we can assume that $0$ separates the interiors of $I$ and $J$. Let us assume that $I$ is to the left of zero. We change the problem. Set $\tilde{\sigma}(dx) = \sigma(-dx)$ for $x \geq 0$, and for $f \in L^2(I, \sigma)$, set
\( \phi(x) = f(-x) \). Then,
\[
\langle H_\sigma f, g \rangle_w = \int_{-\infty}^{0} \int_{0}^{\infty} \frac{f(y)g(x)}{y-x} \sigma(dy)w(dx)
\]
(6.15)
\[
= -\int_{0}^{\infty} \int_{-\infty}^{0} \frac{\phi(y)g(x)}{x+y} \bar{\sigma}(dy)w(dx).
\]

The double integral is split into dual terms, one of which is

(6.16)
\[
\int_{0}^{\infty} \int_{-\infty}^{0} \frac{\phi(y)g(x)}{x+y} \bar{\sigma}(dy)w(dx).
\]

We analyze this bilinear form.

Note that \( x + y \simeq x \) in (6.16). Thus, it suffices to estimate
\[
\int_{0}^{\infty} \int_{0}^{x} \left| \frac{\phi(y)}{x} \bar{\sigma}(dy) \right|^2 w(dx) = \int_{0}^{\infty} \int_{0}^{x} \left| \frac{\phi(y)}{x} \bar{\sigma}(dy) \right|^2 \frac{w(dx)}{x^2} \leq B^2 \| \phi \|_2^2.
\]

where \( B \) is as in (6.12), and \( \hat{w}(dx) = \frac{w(dx)}{x^2} \) and \( \sigma = \bar{\sigma} \).

It remains to estimate the constant \( B \). For any \( 0 < r < \infty \),
\[
\int_{r}^{\infty} \bar{\sigma}(dy) \int_{r}^{\infty} \frac{\sigma(-r,0)}{r} \hat{w}(dx) = \frac{\sigma(-r,0)}{r} \int_{r}^{\infty} \hat{w}(dx) \leq A_2.
\]

The more precise conclusion (6.9) can be read off from this inequality. Recall that (6.15) is split into two bilinear forms, and we have only considered one of them. This explains the symmetric form of (6.9). □

6.2. The Different Subcases of Lemma 4.3. Lemma 4.3 follows from appropriate bounds on these bilinear forms, and their duals.

\[
B^{\text{nearby}}(f,g) := \sum_{I,J : 2^{-r} |I| \leq |J| \leq |I|} |\langle H_\sigma \Delta_I^r f, \Delta_J^w \phi \rangle_w|,
\]

\[
B^{\text{far}}(f,g) := \sum_{I,J : 2^{-r} |J| \leq |I|} |\langle H_\sigma \Delta_I^r f, \Delta_J^w \phi \rangle_w|,
\]

\[
B^{\text{close}}(f,g) := \sum_{I,J : 2^{-r} |I| \leq |J| \leq 3|I|} |\langle H_\sigma \Delta_I^r f, \Delta_J^w \phi \rangle_w|,
\]

\[
B^{\text{adjacent}}(f,g) := \sum_{I,J : |I| \leq |J|} |\langle H_\sigma \Delta_I^r f, \Delta_J^w \phi \rangle_w|.
\]

Here, we are suppressing the role of the truncations, since their role is already indicated by the lemmas above.

Lemma 6.17. For \( * \in \{ \text{nearby, far, close, adjacent} \} \), there holds
\[
B^*(f,g) \leq A_2^{1/2} \| f \|_w \| g \|_w.
\]
6.3. **The Nearby Term.** One can check directly that for each interval $I$, with child $I'$, there holds $|E^\sigma_{I'} h^\sigma_{I'}| \leq \sigma(I')^{-1/2}$. It then follows from (6.8) that $|\langle H_{\sigma} h^\sigma_I, h^\sigma_{I'} \rangle_w| \leq \mathcal{H}$. And then,

$$B_{\text{nearby}}(f, g) \leq \mathcal{H} \sum_{I, J: 2^{-r}|I| \leq |J| \leq |I|} |\hat{f}(I)| |\hat{g}(J)| \leq \mathcal{H} \|f\|_\sigma \|g\|_w.$$ 

The last line follows from the fact that for each $I$, there are only a bounded number of $J$ occurring in the sum.

Here, and below, we will be using the notation $\hat{f}(I) = \langle f, h^\sigma_I \rangle$. 

6.4. **The Far Term.** For integers $s \geq r$, the sum below specifies a relative length for the interval $J$ with respect to $I$.

$$\sum_{I, J: 2^{|J|}=|I|} |\langle H_{\sigma} \Delta^\sigma_I f, \Delta^\sigma_J \phi \rangle_w| \leq \sum_{I, J: 2^{|J|}=|I|} |\hat{f}(I)| |\hat{g}(J)| \frac{\sqrt{\sigma(I)w(J)}}{|\langle J \rangle + \text{dist}(I, J)|^2}$$

$$\leq \sum_{I} |\hat{f}(I)| \sqrt{\sigma(I)} \sum_{J: 2^{|J|}=|I|} |\hat{g}(J)| \frac{\sqrt{w(J)} |\langle J \rangle|}{|\langle J \rangle + \text{dist}(I, J)|^2}$$

Using Cauchy–Schwarz in $I$, we clearly pick up term $\|f\|_\sigma$. The square of the remaining term is

$$\sum_I |\hat{f}(I)|^2 \sum_{J: 2^{|J|}=|I|} |\hat{g}(J)|^2 \frac{\sigma(I)w(J)}{|\langle J \rangle| + \text{dist}(I, J)|^2}^2$$

$$\leq \sum_I |\hat{f}(I)|^2 \sum_{J: 2^{|J|}=|I|} |\hat{g}(J)|^2 \frac{\sigma(I)w(J)}{|\langle J \rangle| + \text{dist}(I, J)|^2} \sum_{J: 2^{|J|}=|I|} |\hat{g}(J)|^2 \frac{\sigma(I)w(J)}{|\langle J \rangle| + \text{dist}(I, J)|^2}$$

$$\leq A_2 \sum_I |\hat{f}(I)|^2 \sum_{J: 2^{|J|}=|I|} |\hat{g}(J)|^2 \frac{\sigma(I)w(J)}{|\langle J \rangle| + \text{dist}(I, J)|^2} \leq 2^{-2s} A_2 \|g\|_w^2.$$ 

The second line follows from (6.4), and then using Cauchy-Schwarz, so that one can appeal to the $A_2$ condition. The last line follows by inspection. This estimate is summable in $s > r$, so this case is complete.

6.5. **The Close Term.** For integers $s \geq r$, the sum below a relative length of $J$ with respect to $I$. Applying (6.6),

$$\sum_{I, J: 2^{|J|}=|I|} |\langle H_{\sigma} \Delta^\sigma_I f, \Delta^\sigma_J \phi \rangle_w| \leq 2^{1-2s} \sum_{I, J: 2^{|J|}=|I|} |\hat{f}(I)| |\hat{g}(J)| \frac{\sqrt{\sigma(I)w(J)}}{|I|}$$

$$\leq 2^{1-2s} \sum_I |\hat{f}(I)| \sqrt{\sigma(I)} \sum_{J: 2^{|J|}=|I|} |\hat{g}(J)| \sqrt{w(J)}$$
We have the geometric decay in $s$. Apply Cauchy–Schwarz, one term is $\|f\|_w$. The other term, squared, is
\[
\sum_1 \frac{\sigma(I)}{|I|^2} \sum_{J: |2^*I| = |I|} \hat{g}(J)^2 \times \sum_{J: |2^*I| = |I|} \frac{w(J)}{n} \leq A_2 \sum_1 \frac{\hat{g}(J)^2}{|I|} \leq A_2 \|g\|_w^2.
\]

This completes the estimate.

6.6. **The Adjacent Term.** We argue as in the previous case. It is easy to see that $|\mathbb{E}_{I-1} f| \leq |\hat{f}(1)| - (I - I_f)^{-1/2}$.

For $\theta \neq \theta' \in \{\pm\}$, and consider the sum below, where $s$ plays the same role as before.
\[
\sum_{I,J: |2^*I| = |I|, \ J \subset \pi^I \theta} |\mathbb{E}_{I}^\sigma \Delta_I^s f \cdot \langle H_{\sigma I_0}, \Delta_J^w g \rangle_w| \leq \sum_{I,J: |2^*I| = |I|, \ J \subset \pi^I \theta} |\hat{f}(1)\hat{g}(J)| \frac{\sqrt{|\sigma(I_0)|} w(J)}{|I|} \leq A_2 \|f\|_w \|g\|_w.
\]

The details are suppressed.

6.7. **Context and Discussion.** The techniques of this section are all drawn from the work of Nazarov-Treil-Volberg [35, 58], aside from the use of the two weight Hardy inequality, which is drawn from [20].

7. **Functional Energy Inequality**

We state an important multi-scale extension of the energy inequality (3.18).

**Definition 7.1.** Fix $s \in (1, 2)$. Let $\mathcal{F}$ be a collection of dyadic intervals. A collection of functions $\{g_\mathcal{F}\}_{\mathcal{F} \in \mathcal{F}}$ in $L^2(w)$ is said to be $\mathcal{F}_e$-adapted if the Haar support $\mathcal{J}(\mathcal{F})$ of $g_\mathcal{F}$ is contained in $\{J : \pi^J_\mathcal{F} = \mathcal{F}, \ J \in \mathcal{F}\}$. Here we set $\pi^J_\mathcal{F} = \pi^J_\mathcal{F}$, and $\pi^J_\mathcal{F}$ is the smallest member of $\mathcal{F}$ that strictly contains $\pi^J_\mathcal{F}$ (that is, it is the $\mathcal{F}$-grandparent of $J$).

The main result of this section is this extension of the energy inequality (3.18).

**Theorem 7.2.** The inequality below holds for all non-negative $h \in L^2(\sigma)$, all $\sigma$-Carleson collections $\mathcal{F}$, that is satisfying (4.7), and all $\mathcal{F}_e$-adapted collections $\{g_\mathcal{F}\}_{\mathcal{F} \in \mathcal{F}}$:
\[
\sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(\mathcal{F})} \mathcal{P}(w \sigma, J^*) \langle \left\langle h, J^* \right\rangle_w , \ g_\mathcal{F} J^* \rangle_{w} \leq \mathcal{H}(\|h\|_w \left[ \sum_{F \in \mathcal{F}} \|g_\mathcal{F}\|_w^2 \right]^{1/2}.
\]

Here $\mathcal{J}^*(\mathcal{F})$ consists of the maximal intervals $J$ in the collection $\mathcal{J}(\mathcal{F})$. The inequality above should be thought of as a two weight inequality for the Poisson integral, a decisive step, since there is a two weight inequality for the Poisson operator proved by Sawyer [50], which is recalled in §8. It reduces the full norm inequality above to simpler testing conditions. The testing conditions are expressed in terms of the Poisson integral, many of which are controlled by the $A_2$ condition. There is one that requires a recursive application of the energy inequality.
7.1. Two Weight Poisson Inequality. Consider the weight $\mu$ on $\mathbb{R}^2_+$ given by

$$
\mu \equiv \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F)} \left\| P^w_{F,J} \mathbf{x} \right\|^2 \delta_{(x_J,J)}.
$$

Here, $P^w_{F,J} := \sum_{J' \in \mathcal{J}(F) : J' \subset J} \Delta^w_{J}$. We can replace $x$ by $x - c$ for any choice of $c$ we wish; the projection is unchanged. And $\delta_q$ 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:

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

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

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

where $x_J$ is the center of the interval $J$. Theorem 7.2 then follows by duality.

Phrasing things in this way brings a significant advantage: The characterization of the two-weight inequality for the Poisson operator Theorem G reduces the full norm inequality above to these testing inequalities: For any (non-dyadic) interval $I$,

$$
\int_{\hat{I}} P(\sigma \cdot I)^2 d\mu(x,t) \lesssim 9\ell^2 \sigma(I), \quad (7.4)
$$

and

$$
\int_{\hat{I}} P^*(t\hat{I}\mu)^2 d\mu dx dt \lesssim A_2 \int_{\hat{I}} t^2 d\mu(dx dt), \quad (7.5)
$$

where $\hat{I} = I \times [0,|I|]$ is the Carleson box over $I$ in the upper half-plane. The dual Poisson operator $P^*(t\hat{I}\mu)$ is

$$
P^*(t\hat{I}\mu)(x) = \int_0^t \frac{t^2}{t^2 + |x - y|^2} \mu(dy dt).
$$

7.2. The Poisson Testing Inequality: The Core. We prove an essential subcase (7.4), by a recursion on the energy inequality, namely the the integral below is restricted to $\hat{I}$:

$$
\int_{\hat{I}} P(\sigma \cdot I)^2(x,t)^2 \mu(dx dt) \lesssim 3\ell^2 \sigma(I).
$$

As the energy inequality is the main tool, let us recall that the constant in that inequality only depends upon the $A_2$ and one of the testing inequalities.

Since $(x_J,J) \in \hat{I}$ if and only if $J \subset I$ (since $I$ is dyadic), we have

$$
\int_{\hat{I}} P(\sigma \cdot I)(x,t)^2 \mu(dx dt) = \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F) : J \subset I} P(\sigma \cdot I)(x_J,J)^2 \left\| P^w_{F,J} \mathbf{x} \right\|^2
$$

$$
\lesssim \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F) : J \subset I} P(\sigma \cdot I,J)^2 \left\| P^w_{F,J} \mathbf{x} \right\|^2. \quad (7.6)
$$
We dispense with an easy case. Let $J^2$ be those intervals $J \in \mathcal{J}^*(F)$, with $J \subset I$ and $I \subset F$. These intervals are pairwise disjoint, contained in $I$, and hence by the energy inequality (3.18),

\begin{equation}
(7.7) \sum_{J \in J^2} P(\sigma \cdot I, J)^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2 \leq 3\varepsilon^2 \sigma(I).
\end{equation}

We will use these intervals to set up a recursion: $\mathcal{F}^*$ are the maximal intervals $F^* \in \mathcal{F}$ such that $F^* \subset I$. (The $\mathcal{F}$-children of $I$.) We show this estimate, with holes in the argument of the Poisson operator:

\begin{equation}
(7.8) \sum_{F^* \in \mathcal{F}^*} \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F): J \subset I} P(\sigma \cdot (I - F^*), J)^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2 \leq 3\varepsilon^2 \sigma(I).
\end{equation}

This estimate can be recursively applied, namely note that

\begin{equation}
(7.6) \leq (7.7) + \text{LHS}(7.8) + \sum_{F^* \in \mathcal{F}^*} \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F): J \subset I} P(\sigma \cdot F^*, J)^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2.
\end{equation}

But, then this same estimate applies to the terms in this last sum, and the collection of intervals $\mathcal{F}$ is $\sigma$-Carleson, so we see that (7.4) holds.

It remains to prove (7.8). Fix data $F^* \in \mathcal{F}^*$, interval $J^* \in \mathcal{J}^*(F^*)$, interval $F \in \mathcal{F}$ with $F \subset I$, and interval $J \in \mathcal{J}^*(F)$. Then, $J \subset J^* \subset F^*$. And, invoking goodness of the interval $J$ allows this ‘change of measure’ argument. In the middle line, we use (5.18).

\begin{align*}
P(\sigma \cdot (I - F^*), J)^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2 &= \left[ \int_{I - F^*} \frac{|J|}{|J|^2 + \text{dist}(x, J)^2} \sigma(dx) \right]^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2 \\
&\leq \left[ \int_{I - F^*} \frac{|J^*|}{\text{dist}(x, J^*)^2} \sigma(dx) \right]^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2 \\
&\leq \left[ \int_{I - F^*} \frac{|J^*|}{|J^*|^2 + \text{dist}(x, J^*)^2} \sigma(dx) \right]^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2.
\end{align*}

This just depends upon the fact that the distance from $J$ to $I - F^*$ is a large multiple of $|J|$. Therefore, we have that

\begin{equation}
\text{LHS}(7.8) \leq \sum_{F^* \in \mathcal{F}^*} \sum_{J \in \mathcal{J}^*(F^*)} P(\sigma \cdot I, J)^2 \sum_{J: J \subset I} \left( \frac{x}{|J^*|}, h_{J^*}^w \right)^2 \|P_{F,J}^w \frac{x}{|J|}\|_w^2 \leq 3\varepsilon^2 \sigma(I).
\end{equation}

That is, we see that the term in (7.8) is dominated by an instance of the energy inequality.

7.3. **The Poisson Testing Inequality: The Remainder.** Now we turn to proving the following estimate for the remainder of the first testing condition (7.4):

\begin{equation}
\int_{3I - I} P(\sigma \cdot I)^2 d\mu \leq A_2 \sigma(I).
\end{equation}

With $F_I$ the unique $F \in \mathcal{F}$ with $J \in \mathcal{J}^*(F)$, estimate

\begin{equation}
\int_{3I - I} P(\sigma \cdot I)^2 d\mu = \sum_{J: (x_J, |J|) \in 3I - I} P(\sigma \cdot I(x_J, |J|))^2 \|P_{F_I,J}^w \frac{x}{|J|}\|_w^2
\end{equation}
organize the sum according to the length of $J$ and then use the Poisson inequality (6.2), available because of goodness of intervals $J$.

\[
\leq \sum_{n=0}^{\infty} \sum_{|J| \geq 3 |J|} 2^{-n(2-4\epsilon)} \frac{\sigma(I)^2}{|I|^2} w(J)
\]

\[
\leq \frac{\sigma(3I) w(3I)}{|3I|^2} \sigma(I) \sum_{n=0}^{\infty} 2^{-n(2-4\epsilon)} \leq A_2 \sigma(I).
\]

### 7.4. The Dual Poisson Testing Inequality

We are considering (7.5). Note that the expressions on the two sides of this inequality are

\[
\int_{I} t^2 \mu(dx dt) = \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}_F(F)} \|P_{t,J} x\|_w^2,
\]

\[
\mathbb{P}^* (\hat{t} \mu)(x) = \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}_F(F)} \frac{\|P_{t,J} x\|_w^2}{|J|^2 + |x - x_j|^2}.
\]

Expand the square in $\|\mathbb{P}^* (\hat{t} \mu) \cdot I\|_0^2$, writing it as a sum over $s \geq 0$ of the terms $T_s$ defined below. In the sum, the relative lengths of $J$ and $J'$ are fixed by $s$.

\[
T_s := \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}_F(F)} \sum_{J' \in \mathcal{J}_F(F')} \int_{|J'| = 2^{-s}|J|} \frac{\|P_{t,J} x\|_w^2}{|J|^2 + |x - x_j|^2} \cdot \frac{\|P_{t,J'} x\|_w^2}{|J'|^2 + |x - x_{j'}|^2} d\sigma
\]

\[
\leq M_s \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}_F(F)} \sum_{J' \in \mathcal{J}_F(F')} \int_{|J'| = 2^{-s}|J|} \frac{1}{|J|^2 + |x - x_j|^2} \cdot \frac{\|P_{t,J'} x\|_w^2}{|J'|^2 + |x - x_{j'}|^2} d\sigma.
\]

where $M_s \equiv \sup_{F \in \mathcal{F}} \sup_{J \in \mathcal{J}_F(F)} \sum_{J' \in \mathcal{J}_F(F') \setminus \text{dist}(J, J') < (n + 1)|J|}$.

The term $M_s$ is at most a constant times $A_2 2^{-s}$, which is then trivially summable in $s \geq 0$. This argument depends upon a case analysis. Fix $J$ as in the definition of $M_s$, and estimate $\|P_{t,J} x\|_w^2 \leq w(J') |J|^2$. Let us consider $J'$ with $n|J| \leq \text{dist}(J, J') < (n + 1)|J|$. In the case of $n = 0, 1, 2$, estimate

\[
\frac{w(J')}{|J'|} \int \frac{|J'|^2}{|J|^2 + |x - x_j|^2} \cdot \frac{|J'|}{|J'|^2 + |x - x_{j'}|^2} d\sigma \leq A_2 2^{-s},
\]

This estimate has to be summed over $J' \subset J$, leading to a final estimate of $2^{-s} A_2$. 

\[
= \sum_{J : J \subset 3I - 1} \mathbb{P}(\sigma \cdot I)(x_j, |J|)^2 \|P_{t,J} x\|_w^2
\]
In the case that \( \sigma \geq 3 \), further require that \( J' \) be to the right of \( J \), and set \( t_n = x_1 + \frac{1}{2}|J| \). The case of \( J' \) being to the left of \( J \) is entirely similar. Note the restriction on the range of integration below, which is the easy case. The first term in the integral is dominated by its \( L^\infty \) norm.

\[
\sum_{F \in \mathcal{F}} \sum_{J' \in \mathcal{J}^+(F)} \sum_{n(\|J'\| \leq \text{dist}(J,J') < (n+1)|J|) \cap \mathbb{Z}} \frac{w(J')}{|J'|} \int_{I \cap [t_n, \infty)} \frac{|J'|^2}{|J|^2 + |x - x_j|^2} \frac{|J'|}{|J'|^2 + |x - x_{J'}|^2} \, d\sigma
\leq 2^{-2s} n^{-2} \sum_{F \in \mathcal{F}} \sum_{J' \in \mathcal{J}^+(F)} \sum_{n(\|J'\| \leq \text{dist}(J,J') < (n+1)|J|) \cap \mathbb{Z}} \frac{w(J')}{|J'|} \int_{I} \frac{|J'|^2}{|J'|^2 + |x - x_j|^2} \, d\sigma
\leq 2^{-2s} n^{-2} A_2 \sum_{F \in \mathcal{F}} \sum_{J' \in \mathcal{J}^+(F)} \frac{1}{|J'|} \sum_{n(\|J'\| \leq \text{dist}(J,J') < (n+1)|J|) \cap \mathbb{Z}} 1
\leq 2^{-2s} n^{-2} A_2 .
\]

The last line follows from the fact that each \( J' \) can be in at most one collection \( \mathcal{J}^+(F) \). And this is trivially summable in \( n \geq 3 \) to the claimed bound.

In the complementary range of integration, we also sum in \( n \), and use the \( L^\infty \) norm on the second term in the integral.

\[
\sum_{F \in \mathcal{F}} \sum_{n=1}^{\infty} \sum_{J' \in \mathcal{J}^+(F)} \sum_{n(\|J'\| \leq \text{dist}(J,J') < (n+1)|J|) \cap \mathbb{Z}} \frac{w(J')}{|J'|} \int_{I \cap [-\infty, t_n]} \frac{|J|}{|J|^2 + |x - x_j|^2} \frac{|J'|^2}{|J'|^2 + |x - x_{J'}|^2} \, d\sigma
\leq 2^{-2s} \sum_{n=1}^{\infty} \frac{w(J+n|J|)}{n^2 |J|} \int_{I} \frac{|J|}{|J|^2 + |x - x_j|^2} \, d\sigma
\leq 2^{-2s} P(w,J)P(\sigma,J) \leq 2^{-2s} A_2 .
\]

This is the only place in which the two-sided Poisson \( A_2 \) condition is needed.

8. Two Weight Poisson Inequality

We give a proof of a two weight inequality for the Poisson integral. Let us phrase the inequality. For \( \sigma \) be a weight on \( \mathbb{R} \), and \( \mu \) be a weight on \( \mathbb{R}^2_+ \), consider the inequality

\[
\|P_\sigma f(x, t)\|_{L^2(\mathbb{R}^2_+, \mu)} \leq N_\sigma \|f\|_{L^2(\mathbb{R}, \sigma)}
\]

\[
P_\sigma f(x, t) := \int_{\mathbb{R}} f(y) \cdot \frac{t}{(t + |x - y|^2)^2} \, d\sigma(y) := p_t * \nu(x) .
\]

**Theorem G.** The following conditions are equivalent

(1) The two weight inequality (8.1) holds
(2) The testing inequalities below hold uniformly over all intervals \( I \).

\[
\int_{I} (P_\sigma I)^2 \, d\mu(x, t) \leq T_\sigma^2 I(\sigma),
\]

(8.1)
\[ \int_{I} P^*_\mu(t\hat{I})^2\ d\sigma(x) \leq T^2_{\tilde{P}} \int_{I} t^2\mu(dx\ dt). \]

In the second line, \( P^* \) is the dual operator, mapping \( L^2(\mathbb{R}^2_+, \mu) \) to \( L^2(\mathbb{R}, \sigma) \), and \( \hat{I} := I \times [0, |I|] \) is a Carleson cube.

(3) These inequalities hold uniformly over all intervals \( I \) in a fixed dyadic grid \( \mathcal{D} \).

\[ \int_{3I} P_\sigma(\cdot | I)^2\ d\mu(x, t) \leq T^2_{\tilde{P}, \sigma} I, \quad 3\hat{I} := (3I) \times [0, |I|], \]
\[ \int_{3I} P^*_\mu(t\hat{I})^2\ d\sigma(x) \leq T^2_{\tilde{P}, \mu} \int_{I} t^2\mu(dx\ dt). \]

Moreover, for the best constants above, \( N_\sigma \simeq T_\sigma \simeq T_{\sigma, \mu} \).

The difference between the two sets of testing conditions above is subtle, but important for us. The particular measure \( \mu \) that we are interested in is constructed from a (random) dyadic grid, and that structure makes verification of the second set of testing conditions more convenient for us.

As it turns out, it is the equivalence between the norm condition and the dyadic testing conditions was first proved by Sawyer [50]. The equivalence between the testing conditions, uniformly over all intervals, and the norm inequality has recently been greatly simplified. Roughly speaking the difference between the two testing conditions is mediated by an \( A_2 \) condition. But, we will prove that the testing conditions, individually imply the norm condition.

First the testing conditions over all intervals. There is a more general dyadic variant. Fix a choice of dyadic grid \( \mathcal{D} \), and let \( \{\lambda_I : I \in \mathcal{D}\} \) be non-negative constants, with which we define an operator

\[ Tf := \sum_{I \in \mathcal{D}} \lambda_I \int_{I} f\ dx \cdot 1_{\hat{I}}. \]

**Theorem H.** Let \( \sigma \) be a weight on \( \mathbb{R} \) and \( \mu \) a weight on \( \mathbb{R}^2_+ \). The inequality \( \|T_\sigma f\|_{L^2(\mathbb{R}^2_+, \mu)} \leq N_\sigma \|f\|_\sigma \) if and only if these testing inequalities hold: For a choice of constant \( T \), uniformly over all intervals \( I \),

\[ \int_{I} (T_\sigma I)^2\ d\mu \leq T\sigma(I), \quad \int_{I} (T^*_\mu \hat{I})^2\ d\sigma \leq T\mu(\hat{I}). \]

Moreover, the best constant \( N_\sigma \) is comparable to \( T \).

We will deduce the Poisson inequality from this dyadic theorem, and then prove the dyadic result. There is no fundamentally Hilbertian aspect to the proof, so the \( L^p \) variant of the theorem holds. Though it would be easy to include these details, we don’t do so.

8.1. **Proof of Theorem G: Parallel Corona.** The Poisson average is not directly comparable to dyadic operators, leading to this modification of the operator. Set

\[ \tilde{P} f(x, t) = \int \frac{f(y)}{t^2 + (x - y)^2}\ dy. \]

Comparing to the Poisson, we have set the numerator in \( \frac{1}{t^2 + |x-y|^2} \) to 1. Note that in the two weight setting, the factor \( y \) can be absorbed into the weight. To deduce the two weight theorem for the Poisson, it is equivalent to deduce the two weight theorem for \( \tilde{P} \).
The operator \( \tilde{P} \) is an average of dyadic operators given by
\[
T_D f := \sum_{I \in \mathcal{D}} |I|^{-2} \int_I f(x) \, dx \cdot 1_I.
\]

**Lemma 8.3.** There holds for all non-negative \( f \),
\[
\mathbb{E}_D T_D f \simeq \tilde{P} f,
\]
where the expectation on the left is over the random grids as defined in §2.3.

**Proof.** For any dyadic grid, the kernel of \( T_D \) is
\[
\sum_{l \in \mathcal{D}} \frac{1}{|I|^2} 1_I(y) 1_I(x,t) \leq \frac{1}{t^2 + (x-y)^2}.
\]
So, it only remains to prove the reverse inequality, on average. For any point \( x \in \mathbb{R} \), with the selection of random grids, the chance that \( x \) is in the middle third of an interval \( I \in \mathcal{D} \), given that \( x \in I \), is \( \frac{1}{3} \):
\[
\mathbb{P}(x \in \frac{1}{3}I : x \in I) = \frac{1}{3}, \quad x \in \mathbb{R}.
\]
Thus, if \( t > 0 \), and \( |x-y| > t \), consider \( I = I(x,y,t) \in \mathcal{D} \) with \( 8|x-y| \geq |I| \geq 4|x-y| \). With probability \( \frac{1}{3} \), we will have \( y \in I \), hence
\[
\mathbb{E}_I \frac{1}{|I|^2} 1_I(y) 1_I(x,t) \gtrsim \frac{1}{t^2 + (x-y)^2}.
\]
The same inequality with \( |x-y| \leq t \) is trivial, with a selection of two intervals I, so the proof is complete. \( \Box \)

Thus, assuming Theorem H, we see that the hypotheses
\[
\int_1 (\tilde{P}_t I)^2 d\mu \leq \sigma(I), \quad \text{and} \quad \int_1 (\tilde{P}_t \tilde{I})^2 d\sigma \leq \mu(\tilde{I})
\]
imply the testing conditions (8.2) for the dyadic operator, uniformly over the choice of dyadic grid. Hence, the dyadic operators are uniformly norm bounded, and thus, so is \( \tilde{P} \).

### 8.2. Proof of Theorem H: The dyadic approach.

We bound the bilinear form \( \langle T_\sigma f, g \rangle_{\mu} \), assuming that \( f \in L^2(\sigma) \) and \( g \in L^2(\mathbb{R}^d, \mu) \) are non-negative, \( f \) supported on an interval \( I_0 \), and \( g \) supported on \( \hat{I}_0 \). The main tool is the (ordinary) Calderón-Zygmund stopping data for \( f \) and \( g \). Let \( \mathcal{F} \) be the stopping cubes for \( f \), thus \( I_0 \in \mathcal{F} \), and given that \( F \in \mathcal{F} \), the \( \mathcal{F} \)-children are the collections
\[
\text{ch}_\mathcal{F}(F) := \{ F' \text{ is a maximal dyadic subinterval of } F \text{ such that } \mathbb{E}^\sigma_F F' > 10 \mathbb{E}^\sigma_F f \}.
\]
Define \( f_F := \mathbb{E}^\sigma_F f \cdot \mathbb{E}^\sigma_f + \sum_{F' \in \text{ch}_\mathcal{F}(F)} \mathbb{E}^\sigma_f f \cdot F' \). There holds \( \mathbb{E}^\sigma_f f = \mathbb{E}^\sigma_f f_{\pi F} \). And, by construction,
\[
\sum_{F \in \mathcal{F}} f_F^2 \leq 4 \| \mathbb{E}^\sigma_f f \cdot F \|_2^2 \leq \sum_{F \in \mathcal{F}} (\mathbb{E}^\sigma_f f)^2 \sigma(F) \leq \| f \|_\sigma^2.
\]
The second inequality follows since the averages increase at least geometrically.

Use the similar notation \( G \), and \( \hat{G} \) for the stopping data for \( g \). Note that we only compute averages of \( g \) on Carleson cubes, so that \( G \) can be taken to be intervals. For an interval \( I \), define \( \pi I = (\pi_F I, \pi_G I) \).
Expand the inner product as
\[(T_\sigma f, g)_\mu = \sum_{I \in D} \lambda_I \sum_{F \in F} \sum_{G \in G} E^\mu_T f \cdot E^\mu_Q g \cdot \sigma(I) \mu(\hat{I})
= \sum_{F \in F} \sum_{G \in G} \lambda_I E^\mu_T f \cdot E^\mu_Q g \cdot \sigma(I) \mu(\hat{I}).\]

The condition \(\pi I = (F, G)\) implies that \(F \cap G \neq \emptyset\), and below we will concentrate on the case of \(I \subset G \subset F\), deducing this case from the first testing condition in (8.2). The case of \(I \subset F \subset G\) is handled by duality.

Note that if \(\pi_F G \subset F\), we contradict \(\pi_F I = F\), hence we can further restrict to the case of \(\pi_F G = F\). Observe that, holding \(F \in F\) fixed, and using \(E^\mu_T f \leq E^\mu_T f\), there holds
\[E^\mu_T f \sum_{G \in G} \lambda_I E^\mu_Q g \cdot \sigma(I) \mu(\hat{I})
\leq E^\mu_T f \int_{F} T_\sigma F \cdot \sum_{G \in G} g_G \, d\mu
\leq T E^\mu_T f \cdot \sigma(F)^{1/2} \left\| \sum_{G \in G : \pi_F G = F} g_G \right\|_{L^2(\mathbb{R}^n, \mu)}.
\]

The point is that, using positivity, the testing inequality for \(T_\sigma\) in the middle line, and moreover, the dual function to \(T_\sigma F\) is succinctly described in terms of the stopping data for \(g\). In view of (8.4), a simple modification of the quasi-orthogonality (4.9) inequality implies that
\[\sum_{F \in F} (8.5) \leq T \|f\|_{\sigma} \|g\|_{L^2(\mathbb{R}^n, \mu)}.
\]

This completes the proof.

8.3. Two Weight Poisson Inequality: Level Set Method. We turn to the proof that the dyadic testing conditions (3) of Theorem G imply the norm inequality for the Poisson operator. We are following the lines of Sawyer’s original argument [50]. Below, are going to write \(J_F\) instead of \(J\).

8.3.1. Initial Steps. Assume that \(\sigma\) is restricted to some large dyadic interval \(I_0\), and that \(\mu\) it restricted to \(3I_0\). This is suppressed in the notation, and we return to it at the construction of the principal cubes below.

Take non-negative \(\phi \in L^2(3I_0, \mu)\), and consider the open sets \(\Omega_k := \{x : \phi > 2^k\} \subset \mathbb{R}^n\). Take \(I_k\) to be a Whitney decomposition of the set \(\Omega_k\). Namely, an interval \(I \in D\) is in \(I_k\) if and only if \(I\) is maximal subject to the conditions \(3I \subset \Omega_k\), but \(5I \not\subset \Omega_k\). These collections have these properties.

Disjoint Cover: \(\Omega_k = \bigcup_{I \in I_k} I\), and the intervals \(I \in I_k\) are either equal or disjoint, aside from endpoints.

Whitney Condition: \(3I \subset \Omega_k\), but \(5I \not\subset \Omega_k\).

Nested Property: If \(I \in I_k\) and \(I' \in I_{k'}\) with \(I \subset I'\), then \(2^k > 2^{k'}\).

Bounded Overlaps: For all \(k \in \mathbb{Z}\), \(\sum_{I \in I_k} 3I \leq 4 \Omega_k\).

The bounded overlaps property requires explanation. For fixed \(k \in \mathbb{N}\), suppose there are intervals \(I, I' \in I_k\), with \(|I'| < 8|I|\), but \(3I' \cap 3I \neq \emptyset\). Then, it follows that for one of the two children \(J\) of \(I\), there holds \(5J \subset 5I\), hence \(5J \not\subset \Omega_k\), which contradicts the maximality of \(I\).

This is the important
Lemma 8.6. [Maximum Principle] There is a constant \(C\) so that
\[
\mathbb{P}^*(\phi \cdot (3\hat{I})^c)(x) \leq C2^k, \quad x \in I, \ I \in \mathcal{I}_k, \ k \in \mathbb{Z}.
\]

Proof. For \(z \in 5I \setminus \Omega_k\), and any \(y\) with \((y, t) \notin 3\hat{I}\), there holds
\[
p_t(x - y) \leq C_p(t(3-y)).
\]
Multiply this by \(\phi(y, t) \cdot 3\hat{I}\), and integrate with respect to \(\mu\), to see that
\[
\mathbb{P}^*(\phi \cdot 3\hat{I})(x) < C\mathbb{P}^*(\phi)(z) \leq C2^{k-1}.
\]

\(\Box\)

8.3.2. The First Estimate. For \(0 < \delta < 1\), we will show that
\[
\|\mathbb{P}^*_\mu \phi\|_\sigma \leq \mathcal{F}_\mathbb{P} \|\phi\|_\mu + \delta \|\mathbb{P}^*_\mu \phi\|_\sigma + \text{remainder}.
\]

The middle term can be absorbed into the left hand side, and the remainder, which is further split into two terms, is the core of the argument.

Begin with the familiar formula, below, where \(m\) is an integer related to the maximum principle.
\[
\|\mathbb{P}^*_\mu \phi\|_\sigma^2 = \sum_{k \in \mathbb{Z}} \int_{\Omega_{k+m} \setminus \Omega_{k+m+1}} (\mathbb{P}^*_\mu \phi)^2 \sigma(dx)
\]
\[
\leq \sum_{k \in \mathbb{Z}} 2^{2k} \sigma(\Omega_{k+m} \setminus \Omega_{k+m+1})
\]
\[
= \sum_{k \in \mathbb{Z}} 2^{2k} \sum_{I \in \mathcal{I}_k} \sigma(I \cap (\Omega_{k+m} \setminus \Omega_{k+m+1}))
\]
\[
(8.8) \quad := \sum_{k \in \mathbb{Z}} 2^{2k} \sum_{I \in \mathcal{I}_k} \sigma(F_k(I)).
\]

The notation in the last line reflects the fact that a given interval \(I\) can be in many collections \(\mathcal{I}_k\), which fact we will have to confront below.

Now, the sum below is restricted by \(0 < \delta < 1\):
\[
\sum_{k \in \mathbb{Z}} 2^{2k} \sum_{I \in \mathcal{I}_k} \sigma(F_k(I)) \leq \delta \|\mathbb{P}^*_\mu \phi\|_\sigma^2.
\]

This is the middle term in (8.7).

For \(I \in \mathcal{I}_k\), with \(\sigma(F_k(I)) > \delta \sigma(I)\), for each \(x \in F_k(I)\), it follows from the maximum principle that
\[
\mathbb{P}^*_\mu \phi(x) = \mathbb{P}^*_\mu(\phi \cdot 3\hat{I})(x) + \mathbb{P}^*_\mu(\phi \cdot (3\hat{I})^c)(x)
\]
\[
\geq \mathbb{P}^*_\mu(\phi \cdot 3\hat{I})(x) - C2^k \geq \mathbb{P}^*_\mu(\phi \cdot 3\hat{I})(x),
\]
for \(m \in \mathbb{N}\) sufficiently large, but absolute.

In the estimate below, an integral over \(\mathbb{R}\), by duality is written as an integral over \(\mathbb{R}^2_+\). In the latter space, we set \(\hat{\Omega}_k := \bigcup_{I \in \mathcal{I}_k} \hat{I}\). The latter squares are disjoint.

\[
2^k \leq \frac{1}{\sigma(F_k(I))} \int_{F_k(I)} \mathbb{P}^*_\mu(\phi \cdot 3\hat{I})(x) \sigma(dx)
\]
\[
= \frac{1}{\sigma(F_k(I))} \int_{3\hat{I}} \mathbb{P}_\sigma(1_{F_k(I)}) \cdot \phi \mu(dx \, dt)
\]
\[
= \frac{1}{\sigma(F_k(I))} \int_{3I-\Omega_{k+m+1}} P_{\sigma}(1_{F_k(I)}) \cdot \phi \mu(dx \, dt) \\
+ \frac{1}{\sigma(F_k(I))} \int_{3\hat{I} \cap \Omega_{k+m+1}} P_{\sigma}(1_{F_k(I)}) \cdot \phi \mu(dx \, dt) \\
=: A(k, I) + B(k, I).
\]

The first term is easy. By positivity and testing, and assuming that \(\sigma(F_k(I)) \geq \delta \sigma(I)\),
\[
A(k, I) \leq \delta^{-1} \frac{1}{\sigma(I)} \int_{3\hat{I} - \Omega_{k+m+1}} P_{\sigma}(1_{F_k(I)}) \cdot \phi \mu(dx \, dt) \\
\leq \delta^{-1} \frac{1}{\sigma(I)} \|P_{\sigma}(1)\|_\mu \|1_{3\hat{I} - \Omega_{k+m+1}} \phi\|_\mu \\
\leq \delta^{-1} \mathcal{T}_P \sigma(I)^{-1/2} \|1_{3\hat{I} - \Omega_{k+m+1}} \phi\|_\mu.
\]

Thus, we should estimate
\[
\sum_{k \in \mathbb{N}} \sum_{I \in I_k \atop \sigma(F_k(I)) \geq \delta \sigma(I)} A(k, I)^2 \sigma(F_k(I)) \leq \delta^{-2} \mathcal{T}_P^2 \|1_{3\hat{I} - \Omega_{k+m+1}} \phi\|_\mu^2 \\
\leq \delta^{-2} \mathcal{T}_P^2 \int_{\mathbb{R}^2_+} \phi^2 \sum_{k \in \mathbb{N}} \sum_{I \in I_k} 1_{3\hat{I} - \Omega_{k+m+1}} \mu(dx \, dt) \\
\leq \delta^{-2} \mathcal{T}_P^2 \|\phi\|_\mu^2.
\]

The last line follows from assertion that
\[
\left\| \sum_{k \in \mathbb{N}} \sum_{I \in I_k} 1_{3\hat{I} - \Omega_{k+m+1}} \right\|_\infty \leq 1.
\]

But this is the direct consequence of the bounded overlaps property.

The inequality (8.7) is established. To be explicit, the remainder term is as below; its analysis is the core of the argument.
\[
\sum_{k \in \mathbb{Z}} \sum_{I \in I_k \atop \sigma(F_k(I)) \geq \delta \sigma(I)} B(k, I)^2 \sigma(F_k(I)) \\
\leq \sum_{k \in \mathbb{Z}} \sum_{I \in I_k \atop \sigma(F_k(I)) \geq \delta \sigma(I)} \frac{1}{\sigma(I)} \left[ \int_{3\hat{I} \cap \Omega_{k+m+1}} P_{\sigma}(1_{F_k(I)}) \cdot \phi \mu(dx \, dt) \right]^2
\]

(8.9)

8.3.3. The Principal Cubes and the Second Estimate. The region of integration in (8.9) is \(3\hat{I} \cap \Omega_{k+m+1}\). This is, by the nested property, the union of \(\{J : J \subset 3I, J \in I_{k+m+1}\}\). Since \(3I \cap F_k(I) = \emptyset\), for \((x, t) \in \hat{I}\), there holds
\[
P_{\sigma}(1_{F_k(I)})(x, t) \simeq \frac{t}{|J|} P_{\sigma}(1_{F_k(I)})(x_J, |J|),
\]
where \(x_J\) is the center of \(J\). This is an important observation, which yields
\[
\int_{\hat{I}} P_{\sigma}(1_{F_k(I)}) \cdot \phi \mu(dx \, dt) \simeq P_{\sigma}(1_{F_k(I)})(x_J, |J|) \int_{\hat{I}} \frac{t}{|J|} \cdot \phi \mu(dx \, dt)
\]
I is fixed. The union of these three intervals is \( \sigma \). The first three lines follow by inspection, with the definition of (8.10). The last equality follows by inspection.

Employing this inequality, the term in (8.9) is

\[
\int_{3I \cap \Omega_{k+m+1}} \mathbb{P}_\sigma(1_{F_k(I)}) \cdot \phi \, \mu(dx \, dt) = \sum_{J \in \Omega_{k+m+1}} \int_{I \subset J} \mathbb{P}_\sigma(1_{F_k(I)}) \cdot \phi \, \mu(dx \, dt)
\]

(8.10)

\[
\lesssim \sum_{J \in \Omega_{k+m+1}} \int_{I \subset J} (\mathbb{P}_\sigma I) t^{-1} \bar{\mu}(dx \, dt) \times \frac{1}{\bar{\mu}(J)} \int_J \phi t^{-1} \bar{\mu}(dx \, dt).
\]

This calculation suggests that we use the maximal function

\[
M_{\bar{\mu}} \psi := \sup_{I \in \mathcal{D}} \| \tilde{\psi} \|_{\bar{\mu}}^2
\]

which is a bounded operator on \( L^2(\bar{\mu}) \).

Set the principal cubes, or intervals, as follows. Initialize \( \mathcal{G} \) to be \( \{ I_0 \} \), where this is the large interval on which \( \sigma \) is supported. In the inductive stage, for \( I \in \mathcal{G} \) minimal, add to \( \mathcal{G} \) those maximal dyadic children \( J \) such that \( \alpha(J) := \mathbb{E}_J^\mu \phi t^{-1} \geq 10 \alpha(I) \). If there are no such children, for any minimal interval in \( \mathcal{G} \), the construction stops. The maximal function bounds imply

(8.11)

\[
\sum_{I \in \mathcal{G}} \alpha(I)^2 \bar{\mu}(I) \lesssim \| \phi t^{-1} \|_{\bar{\mu}}^2 = \| \phi \|_{\bar{\mu}}^2.
\]

The last equality follows by inspection.

The estimate below controls part of the sum in (8.9). Appealing to (8.10), but also imposing a condition on the principal cubes, estimate as below. In the sum over \( \mathcal{I}_{k+m+1} \),

\( I_0 \in \{ I, I+|I|, I-|I| \} \)

is fixed. The union of these three intervals is \( 3I \). The bound below will be independent of the choice of \( I_0 \).

\[
\sum_{I \in \mathcal{I}_k} \sigma(I)^{-1} \left[ \sum_{J \in \mathcal{I}_{k+m+1}} \int_{I \subset J} (\mathbb{P}_\sigma I) t^{-1} \bar{\mu}(dx \, dt) \times \frac{1}{\bar{\mu}(J)} \int_J \phi t^{-1} \bar{\mu}(dx \, dt) \right]^2
\]

\[
\lesssim \sum_{I \in \mathcal{I}_k} \sigma(I)^{-1} \alpha(\bar{I}_0) \left[ \sum_{J \in \mathcal{I}_{k+m+1}} \int_{I \subset J} (\mathbb{P}_\sigma I) t^{-1} \bar{\mu}(dx \, dt) \right]^2
\]

\[
\lesssim \sum_{I \in \mathcal{I}_k} \sigma(I)^{-1} \alpha(\bar{I}_0) \left[ \int_{I_0} \mathbb{P}_\mu^* (I_0) \sigma(dx) \right]^2
\]
\[ \leq \mathcal{J}_P^2 \sum_{I \in \mathcal{I}_k} \sigma(F(I)) \alpha(\pi G I_0)^2 \hat{\mu}(\hat{I}_0). \]

Duality is used to pass to the testing condition for \( \mathcal{P}^* \).

Sum the term above over \( k \in \mathbb{Z} \).

(8.12) \[ \sum_{k \in \mathbb{Z}} \sum_{I \in \mathcal{I}_k} \alpha(\pi G I_0)^2 \hat{\mu}(\hat{I}_0) = \sum_{G \in \mathcal{G}} \alpha(G)^2 \sum_{k \in \mathbb{Z}} \sum_{I \in \mathcal{I}_k} \hat{\mu}(\hat{I}_0) \]

The difficulty with this sum is that a given interval \( I \) can be in many collections \( \mathcal{I}_k \), so that the term \( \mu(\hat{I}_0) \) can contribute to the sum many times. See however, this Lemma.

**Lemma 8.13.** For any dyadic interval \( I \), the set of integers below consists of at most \( \delta^{-1} \) consecutive integers.

\[ \{ k \in \mathbb{Z} : I \in \mathcal{I}_k, \ \sigma(F(I)) \geq \delta \sigma(I) \} \]

**Proof.** That the integers in the set are consecutive follows from the nested property of the collections \( \mathcal{I}_k \). The sets \( F_k(I) \subset I \), with \( I \) fixed, are pairwise disjoint, as follows from the definition in (8.8). And each has \( \sigma \) measure at least \( \delta \sigma(I) \). \( \square \)

It then follows from the disjoint cover property and (8.11), that

\[ \text{RHS(8.12)} \leq \delta^{-1} \sum_{G \in \mathcal{G}} \alpha(G)^2 \hat{\mu}(\hat{G}) \leq \| \phi \|^2_{\hat{\mu}}. \]

This holds for each of the three possible choice of \( I_0 \), so completes this case.

8.3.4. **The Third Estimate.** It remains to bound the sum over \( k \in \mathbb{Z} \) of the expression below, in which the parents of \( I_0 \) and \( J \) in the collection \( \mathcal{G} \) differ.

\[ \sum_{I \in \mathcal{I}_k} \sigma(I)^{-1} \left[ \sum_{J \subset I, \sigma(F(I)) \geq \delta \sigma(I)} \int_{J} (\mathcal{P}_0 I) t^{-1} \hat{\mu}(\hat{t} dx dt) \right] \]

The term in brackets above is estimated by inserting powers of \( \hat{\mu}(\hat{J})^{\pm 1/2} \) and using Cauchy-Schwarz. This leads to, on the one hand,

\[ \sum_{J \subset I, \sigma(F(I)) \geq \delta \sigma(I)} \hat{\mu}(\hat{J}) \alpha(\pi G J)^2. \]

And, on the other hand, using Cauchy-Schwarz, inspection, and the testing condition,

\[ \sum_{J \subset I, \sigma(F(I)) \geq \delta \sigma(I)} \left[ \int_{J} (\mathcal{P}_0 I) t^{-1} \hat{\mu}(\hat{t} dx dt) \right] \hat{\mu}(\hat{J})^{-1} \leq \sum_{J \subset I, \sigma(F(I)) \geq \delta \sigma(I)} \int_{J} (\mathcal{P}_0 I)^2 \hat{\mu}(\hat{t} dx dt) \]

\[ \leq \mathcal{J}_P^2 \sigma(I). \]

Putting these pieces together, it remains to estimate the sum

(8.14) \[ \sum_{k \in \mathbb{Z}} \sum_{I \in \mathcal{I}_k} \sum_{J \subset I, \sigma(F(I)) \geq \delta \sigma(I), J \subset I, \sigma(F(J)) \geq \delta \sigma(J)} \hat{\mu}(\hat{J}) \alpha(\pi G J)^2 \]
Again, a given interval $J$ can occur many times in the sum above. But, in view of the following Lemma, the sum above is bounded by the expression in (8.11), completing the proof.

**Lemma 8.15.** There is an absolute $C$ so that for any $G \in \mathcal{G}$, the cardinality of the set below is at most $C$.

$$\{ k : \pi_G J = G \, , \, J \in \mathcal{I}_{k+m+1} \text{ contributes to the } k\text{th sum in (8.14)} \}$$

**Proof.** For $G \in \mathcal{G}$, consider data

$$k_1 \geq k_2 \geq \cdots \geq k_n, \quad J_1 \subset J_2 \subset \cdots \subset J_n, \quad \pi_G J_1 = \cdots = \pi_G J_n = G,$$

$$(I_1)_0 \subset (I_2)_0 \subset \cdots \subset (I_n)_0,$$

$$J_t \in \mathcal{I}_{k_t+m+1}, \quad J_t \subset (I_t)_0, \quad I_t \in \mathcal{I}_{k_t}, \quad \pi_G J_t \not\subseteq \pi_G (I_t)_1, \quad 1 \leq t \leq n.$$  

An upper bound on $n$ is what is needed.

All of the intervals $3I_t$ are overlapping. And some of the $k_t$ can be equal, but only a bounded number can be equal, by the bounded overlap property. Hence, after a deleting some data, and relabeling, the $k_t$ can be taken to be strictly decreasing. A number of the $I_t$ can have the same length, but only one of three possible positions, since $J_1 \subset 3I_t$ for all $t$. From Lemma 8.13, there are a bounded number of such $t$ that with $I_t$ having a fixed length and position. Thus, after a further deletion and relabeling, the lengths $|I_t|$ can be taken to be strictly increasing. But then, since $I_t \in \mathcal{I}_{k_t}$ and $J_t \in \mathcal{I}_{k_t+m+1}$, if $m < n$, for some $t$, there holds

$$J_1 \subset (I_t)_0 \subset J_n.$$  

This inclusion contradicts $\pi_G J_1 = \pi_G J_n = G$ and $\pi_G J_1 \not\subseteq \pi_G (I_t)_1$. The Lemma is established. $\square$

8.4. **Context and Discussion.**

8.4.1. The two weight Poisson result is in Sawyer’s paper [50], and the dyadic variant Theorem H was first noted in [18]. Sawyer’s original method of proof, a level set technique, based upon a maximum principle, is followed in [18]. Various extensions of this argument to singular integral settings has been studied in [15, 19], as well as in certain vector valued settings [54]. Sawyer’s paper is largely focused on the two weight inequalities for the fractional integrals, which is widely cited result. The use of the two weight inequality for the Poisson integral herein is the only one I am aware of.

8.4.2. Beginning with Treil [56], there has been a renewed attention on very short proofs of the dyadic positive operator case. The first proof above is based upon the last page of Hytönen [14] (which was in turned suggested by an early approach in [22]), and the approach of Adelman-Pott-Reguera [2]. Also see Tanaka [55].

---

6 Certain combinatorial arguments required by Sawyer’s innovative approach [50] are fully explained in the latter references, such as [18].
9. Compact Operators

We discuss the proof of Theorem 1.7, which revisits much of the proof of the main theorem. An element of that proof was a discussion of the role of truncations. Now, Theorem 1.7 is stated for the ‘hard’ truncations of the introduction, and we readopt the notation $H_{\alpha,\beta}(\sigma f)$ for these ‘hard’ truncations, as defined in (1.2) throughout this section.

We will prove the theorem with these truncations, and then briefly indicate how the same theorem can be obtained for other truncations.

9.1. Necessity. A bounded operator $\Lambda$ on a Hilbert space is compact if and only if for all orthonormal sequences $x_n$, there holds $\lim_n \|\Lambda x_n\| = 0$. Indeed, this condition can be weakened to the following, by a straight forward argument: $\lim_n \|\Lambda x_n\| = 0$ for all sequences $\{x_n\}$ in the Hilbert space, of norm one, which are asymptotically orthogonal, in the sense that $\lim_{t \to \infty} |\langle x_n, x_t \rangle| = 0$, for all $n$.

Assume that $H_{\sigma} : L^2(\sigma) \to L^2(w)$ is bounded and compact, in the sense of (1.1) and Definition 1.6, with the ‘hard’ truncations. By way of contradiction, assume that there is a $\lambda > 0$ and infinite family $\{I_n : n \in \mathbb{N}\}$ of closed intervals that satisfy

\begin{equation}
\int_{I_n} (H_{\sigma} I_n)^2 \, dw \geq \tau \sigma(I_n)
\end{equation}

where $\tau > 0$ is fixed. And, these intervals exhibit the failure of condition (1.12). Thus, for $\Lambda_n \to \infty$, we have $I_n \subset \mathbb{R} \setminus [-\Lambda_n, \Lambda_n]$. Hence, there is a subsequence which is disjoint, hence the normalized indicators are asymptotically orthogonal, in the sense that $\lim_{t \to \infty} |\langle x_n, x_t \rangle| = 0$, for all $n$.

Second, assume that the intervals in (9.1) exhibit the failure of (1.11). Thus, for some $\Lambda > 0$, we have $I_n \subset [-\Lambda, \Lambda]$, and $|I_n| \to 0$. As there are no common point masses, we can assume that the $I_n$ are closed. If there is some infinite subsequence of intervals that are disjoint, we have a contradiction. Thus, there is no such sequence, and $\cap_n I_n = \{x_0\}$. But, $\sigma(I_n) \to 0$, since $\sigma$ is absolutely continuous with respect to Lebesgue measure. Thus, the normalized indicators are asymptotically orthogonal, and we have a contradiction. The duals to these conditions are also clearly necessary.

Now, turn to the necessity of the $\Lambda_2$ conditions (1.8)—(1.10). These are a bit more complicated, since the Poisson averages at interval $I$ can be large due to influences far away from $I$. We need this consequence of our proof of the necessity of the $\Lambda_2$ bound.

**Proposition 9.2.** Assume that $P(\sigma, I) \cdot P(w, I) > \lambda$. Then, there are two compact intervals $I_\sigma$ and $I_w$, and function $f_1$ of $L^2(\sigma)$-norm one supported on $I_\sigma$ such that the distance from $I_\sigma$ and $I_w$ is positive and

$$\lambda \leq \int_{I_w} |H_{\sigma} f_1|^2 \, dw.$$

Moreover, $f_1(x) = c_1 \cdot p_1(x) \cdot I_\sigma(x)$, where $p_1$ is as in (3.2), and $c_1$ is a normalizing constant.

**Proof.** Follow the proof of the necessity of the $\Lambda_2$ condition up to (3.3). From this inequality, for appropriate choice of $\alpha$, we see that there are two complementary half-lines $I_\sigma$ and $I_w$ such that

$$\lambda P(\sigma, I) \leq P(w, I) P(\sigma, I)^2 \leq \int_{I_w} |H_{\sigma}(p_1 \cdot I_\sigma)|^2 \, dw.$$
Here, $p_1$ is as (3.2), and in particular, $\|p_1\|_\sigma \simeq P(\sigma, I)^{1/2}$. This is our conclusion, except that the intervals are infinite and intersect at a single point. But, it is clear that we can make a second approximation, to get the conclusions above. □

We can now turn to the proof of the necessity of the two conditions (1.8)—(1.10). Let $\lambda > 0$, and let $I_n$ be a sequence of intervals such that $P(\sigma, I_n) P(w, I_n) \geq \lambda$ for all $n$. Let $I_{\sigma,n}$ be the intervals given to us by Proposition 3.1; likewise let $f_n$ be the norm one function given to us by this proposition.

Suppose, by way of contradiction, that these intervals exhibit the failure of condition (1.9). Thus, the intervals $I_{\sigma,n} \subset \mathbb{R} \setminus [-\Lambda_n, \Lambda_n]$, where $\Lambda_n \to \infty$. Hence, by compactness of the $I_{\sigma,n}$, we can select a subsequence which are disjoint, hence the $f_n$ are necessarily orthonormal.

The same argument as above works for (1.10) as written. There are three more variants of this condition, symmetric with respect to the roles of $\sigma$ and $w$, or symmetric with respect to the inversion $x \to -x$ on $\mathbb{R}$. They are also necessary for compactness.

It remains to consider the case of the $I_n$ exhibiting the failure of the condition (1.8). Thus, for some $\Lambda > 0$, all the $I_n \subset [-\Lambda, \Lambda]$, and $|I_n| \to 0$. Moreover, by the support condition on the weights imposed in (1.8), we have $I_{\sigma,n} \subset [-\Lambda, \Lambda]$. If some subsequence of the $I_{\sigma,n}$ are disjoint, the functions $f_n$ are orthogonal, hence a contradiction. Thus, we can assume that $\bigcap_n I_{\sigma,n} = \{x_0\}$. One must however note that the condition $|I_n| \to 0$ does not imply anything about $I_{\sigma,n}$. Nevertheless, we now argue that the $(f_n)$ are asymptotically orthogonal:

$$\lim_{n \to \infty} \langle f_n, f_\ell \rangle_\sigma = 0, \quad n \in \mathbb{N},$$

which supplies the required contradiction to compactness.

Since we have very little information about the relative positions of the intervals $I_n$ and those intervals given to us by Proposition 9.2, we should prove that

$$\lim_{n \to \infty} \langle p_{I_n}, p_{I_\ell} \rangle_\sigma = 0, \quad n \in \mathbb{N}.$$ 

Now, assume that $\ell$ is substantially larger than $n$, so that $\tau_\ell^{4/3} := |I_n|/|I_\ell| > 1$. Then, since $\sigma$ does not have point masses, $\sigma(\tau_\ell I_\ell) \to 0$, hence

$$\int_{\tau_\ell I_\ell} p_{I_n}^2 \, d\sigma \to 0.$$

Also, observe, from the definition (3.2), and the condition that $\bigcap_n I_{\sigma, n} = \{x_0\},$

$$\sup_{x \in \tau_\ell I_\ell} \frac{p_{I_n}(x)p_{I_\ell}(x)}{p_{I_n}(x)^2} \lesssim \sup_{x \in \tau_\ell I_\ell} \frac{|I_n|^{1/2} |I_\ell|^{1/2}}{|I_n|^2 + \text{dist}(x, I_n)^2} \cdot \frac{|I_n|^2 + \text{dist}(x, I_n)^2}{|I_n| + \text{dist}(x, I_n)|I_\ell| + \text{dist}(x, I_\ell)}$$

$$\lesssim \frac{|I_\ell|^{1/2}}{|I_n|^{1/2} \tau_\ell |I_\ell|} \sim \tau^{-1} \frac{|I_n|^{1/2}}{|I_\ell|^{1/2}} \sim \tau_\ell^{-1/3}.$$

Therefore,

$$\int_{\mathbb{R} \setminus \tau_\ell I_n} p_{I_n}(x)p_{I_\ell}(x) \, d\sigma(x) \lesssim \tau_\ell^{-1/3} P(\sigma, I_n) \to 0.$$

This completes the proof of the necessity of (1.8).
1.7

conditions for the pair of weights \( (\tilde{w}, \sigma) \) where \( \tilde{w} \) is arbitrary, and \( \sigma \) is continuous, with an absolute modulus of continuity. Indeed, for \( -n < x < x' < n \), and \( x' < x + \alpha \), we have from the definition of the 'hard' truncations (1.2),

\[
|H_{\alpha,\beta}(\sigma f)(x') - H_{\alpha,\beta}(\sigma f)(x)| \leq A + B + C,
\]

Figure 9. The difference \( H_{\alpha,\beta}(\sigma f)(x') - H_{\alpha,\beta}(\sigma f)(x) \) is the integral over six regions, of which three are depicted above: (1) \( x + \alpha \) to \( x' + \alpha \), where there is no cancellation between the two curves; (2) \( x' + \alpha \) to \( x + \beta \), where there is substantial cancellation; and (3) \( x + \beta \) to \( x' + \beta \), where again there is no cancellation. The other three regions are dual, and to the left and below in the diagram above.

9.2. Sufficiency. Note these easy consequences of the sufficient conditions for compactness.

Lemma 9.3. Suppose the pair of weights \( w, \sigma \) meet the sufficient assumptions of compactness of Theorem 1.7. Then, using the notation of that Theorem,

\[
\lim_{n \to \infty} \{\|H_{\alpha,\beta}^{n}_{1}\|_{L^{2}(\sigma_{n}^{1}) \to L^{2}(w_{n}^{1})} + \|H_{\sigma_{n}^{1}}^{1}\|_{L^{2}(\sigma_{n}^{1}) \to L^{2}(w_{n}^{1})} + \|H_{w_{n}^{1}}^{1}\|_{L^{2}(w_{n}^{1}) \to L^{2}(\sigma_{n}^{1})}\} = 0,
\]

\[
\lim_{\lambda \to 0} \sup_{0 < \alpha < \beta < \lambda} \|H_{\alpha,\beta}(\sigma)\|_{L^{2}(\sigma_{n}^{0}) \to L^{2}(w_{n}^{0})} = 0, \quad n > 0.
\]

Note that the role of the truncations is made explicit in the second limit, while it is implicit in the first limit. (\( \sigma = \sigma_{n}^{0} + \sigma_{n}, \) where \( \sigma_{n}^{0} = \sigma 1_{[-n,n]} \)).

Proof. By our main theorem, the norm \( \|H_{\sigma_{n}^{1}}\|_{L^{2}(\sigma_{n}^{1}) \to L^{2}(w_{n}^{1})} \) is bounded above by the \( A_{2} \) and testing conditions for the pair of weights \( (\sigma_{n}^{1}, w_{n}^{1}) \). But the assumptions (1.9), (1.12), and the dual to the latter, show that the \( A_{2} \) and testing constants tend to zero as \( n \) tends to infinity.

The condition on the norm \( \|H_{\sigma_{n}^{1}}\|_{L^{2}(\sigma_{n}^{1}) \to L^{2}(w_{n}^{1})} \) and the third, dual condition, are of a different nature. This time, the weights are supported on disjoint intervals. Hence, the norm of the operator in this case is governed by the Hardy inequality, as we have already detailed in the weak-boundedness inequality. In particular, the inequality (6.9), together with the limit (1.10), and its three variants, show that these last norms also tend to zero. (Indeed, it is this argument that requires this assumption.) Thus, the first limit holds.

The second limit holds, for each fixed \( n > 0 \), since the sufficiency proof shows that the norm of the 'small truncations' are controlled by the 'small' \( A_{2} \) and 'small' testing inequalities. Namely, we have

\[
\sup_{0 < \alpha < \beta < \lambda} \|H_{\alpha,\beta}(\sigma_{t})\|_{L^{2}(\sigma_{n}^{0}) \to L^{2}(w_{n}^{0})} \leq \sup_{|t| \leq n, 0 < \alpha < \beta} \|H_{\alpha,\beta}(\tilde{o}_{t})\|_{L^{2}(\tilde{o}_{n}^{0}) \to L^{2}(\tilde{w}_{n}^{0})}
\]

where \( \tilde{o}_{t} = \sigma \cdot 1_{(1-n(t-n))}. \) This last norm is comparable to the \( A_{2} \) and testing inequalities for the restricted measures \( \tilde{o}_{t} \) and \( \tilde{w}_{n}^{0} \). These tend to zero in \( \lambda \) by (1.8), and (1.11), and its dual. \( \Box \)

It therefore suffices to prove the compactness of \( \{H_{\alpha,\beta}(\sigma_{t}) : \alpha_{0} < \alpha < \beta < 2n\} \), where \( 0 < \alpha_{0} < n \) is arbitrary, and \( \sigma \) and \( w \) are supported on an interval \([-n,n]\). Our observation is that \( H_{\alpha,\beta}(\sigma f)(x) \) is continuous, with an absolute modulus of continuity. Indeed, for \( -n < x < x' < n \), and \( x' < x + \alpha \), we have from the definition of the 'hard' truncations (1.2),

\[|H_{\alpha,\beta}(\sigma f)(x') - H_{\alpha,\beta}(\sigma f)(x)| \leq A + B + C,\]
\[ A = \int_{x+\alpha}^{x'+\alpha} \frac{f(y)}{y-x} \sigma(dy) + \int_{x-\alpha}^{x'-\alpha} \frac{f(y)}{y-x} \sigma(dy) \]
\[ \leq \alpha^{-1} \|f\| \sigma((x-\alpha, x'-\alpha) \cup (x+\alpha, x'+\alpha))^{1/2} \]
\[ B = \int_{x+\alpha}^{x'+\alpha} \frac{x-x'}{(y-x')(y-x')} f(y) \sigma(dy) + \int_{x-\alpha}^{x'-\alpha} \frac{x-x'}{(y-x')(y-x')} f(y) \sigma(dy) \]
\[ \leq \frac{|x-x'|}{\alpha^2} P(\sigma, (x, x+\alpha))^{1/2} \|f\| \sigma, \]
\[ C = \int_{x+\alpha}^{x'+\alpha} \frac{f(y)}{y-x} \sigma(dy) + \int_{x-\alpha}^{x'-\alpha} \frac{f(y)}{y-x} \sigma(dy) \]
\[ \leq \beta^{-1} \|f\| \sigma((x-\beta, x'-\beta) \cup (x+\beta, x'+\beta))^{1/2} \]

This depends upon a routine estimate of the difference. The terms \( A, C \) are associated with the discontinuities of the ‘hard’ truncations, and the middle term \( B \) is associated with the standard smoothness estimate of \( 1/y \). See Figure 9. All estimates depend upon trivial application of Cauchy–Schwarz. And, we only use them for \( \|f\| \sigma \leq 1 \).

Note that since \( \sigma \) does not have point masses, we can make the terms \( B \) and \( C \) uniformly small, as \( \alpha \downarrow 0 \). The Poisson integral term in \( B \) is uniformly bounded over \(-n < x < n\), hence this term, with the leading \( |x-x'| < \alpha \), can also be made uniformly small. It follows that the family of functions

\[ \{H_{\alpha, \beta}(\sigma)(x) : \|f\| \sigma \leq 1, \ \alpha_0 < \alpha < \beta < n\} \]

are pre-compact in \( C[-n, n] \), by Arzela-Ascoli. Hence the family is also compact in the coarser topology of \( L^2(w) \). This completes the proof of compactness of \( H_\sigma \), under the conditions of Theorem 1.7.

9.3. Alternate Truncations. We have considered the ‘hard’ truncations, with the principal points being that one can derive the \( A_2 \) estimate, which in turn depends upon the canonical value of the inner product \( \langle H_\sigma f, g \rangle_w \), as given in (3.7). Since this property holds for a wide set of truncations, as given in (3.6), and we have already seen in Proposition 3.8, that, subject to mild considerations, the norm inequality is independent of how the truncation is taken, the characterization of compactness continues to hold for a wide class of truncations, in particular those given by (3.5) and (3.6).

9.4. Two Examples. We have stated Theorem 1.7 under the assumption that neither weight has a point mass. The case of no common point mass is more complicated, as these two examples show that the ‘vanishing’ criteria can’t characterize compactness in general.

Take \( w = \delta_0 \) be a point mass at zero. Then, take \( \sigma \) to be absolutely continuous with respect to Lebesgue measure, with density equal to

\[ \sigma(x) := \frac{x}{(\log x)^2} 1_{[0, \frac{1}{2}]}(x). \]

Notice that

\[ \int |Hw|^2 \, d\sigma = \int_0^{\frac{1}{2}} \frac{1}{x(\log x)^2} \, dx \simeq 1. \]
And, $L^2(w)$ has but one dimension, thus, $H_w$ is a bounded compact map from $L^2(w)$ to $L^2(\sigma)$. It is clear that as $\lambda \downarrow 0$,
\[
\frac{w([0,\lambda])}{\lambda}P(\sigma,[0,\lambda]) \lesssim \int_0^{1/3} \frac{x}{(\lambda + x)^2(\log x)^2} \; dx \rightarrow \int_0^{1/3} \frac{1}{x(\log x)^2} \; dx \simeq 1.
\]
That is, there is no decay in the $A_2$ ratio.

A second example is to take $w = \delta_0$ as before, and $\sigma(dx) = 1_{(1,\infty)} dx$. Of course the Hilbert transform is bounded for this pair of weights. Note that for any interval $|I| < \frac{1}{4}$ with $\text{dist}(I,0) \leq 2|I|$, we have
\[
P(\sigma,I)P(w,I) \simeq \int_1^\infty \frac{dx}{x^2} \simeq 1.
\]

10. Proof under the Pivotal Assumption

We prove an upper bound for a two weight inequality assuming a pivotal condition on a pair of weights. The set us is as follows. Let $P$ be the best constant in the
\[
\sup_{\alpha, \beta} \left\| \int K_{\alpha,\beta}(x,y) f(y) \sigma(dy) \right\|_w \leq N_T \|f\|_\sigma.
\]
Here, $K_{\alpha,\beta}$ is assumed to satisfy
\[
(\alpha + |x-y|)(|K_{\alpha,\beta}(x,y)| + |x-y| \cdot |\nabla K(x,y)|) \leq C 1_{|x-y| < \beta}
\]
and $K_{\alpha,\beta}(x,y) = K(x-y)$ if $\alpha < |x-y| < \beta$. Then, $T_{\alpha,\beta}(\sigma f)(x) = \int K_{\alpha,\beta}(x,y) f(y) \sigma(dy)$. The $A_2$ condition will be assumed, hence by a variant of Proposition 3.8, the boundedness of these 'smooth' truncations will be equivalent to a wide assortment of truncations.

Let $P$ be the best constant in the pivotal inequality, defined as follows. For any interval $I_0$ and any partition $P$ of $I_0$ into intervals such that neither $\sigma$ nor $w$ have point masses at the endpoints, there holds
\[
(10.1) \sum_{I \in P} P(\sigma,I)^2 w(I) \leq P^2 \sigma(I_0).
\]
We also require that the dual inequality, with the roles of $w$ and $\sigma$ reversed, holds. One can note that this inequality will hold if the maximal function satisfies the two weight inequality $\|M_{\sigma} f\|_w \leq \|f\|_\sigma$, and its dual.

**Theorem 10.2.** [Nazarov-Treil-Volberg [58]] Assume that the pair of weights $w, \sigma$ do not share a common point mass, and satisfy the $A_2$ condition (1.4), and the pivotal conditions hold, namely $P < \infty$. Then, there holds $N_T \leq T_T + A_2^{1/2} + P$, where $T$ is the best constant in the inequalities
\[
\int |T_\sigma I|^2 w(dx) \leq T_T^2 \sigma(I), \quad \int |T_w I|^2 \sigma(dx) \leq T_T^2 w(I).
\]
We give the proof, with the goal of highlighting some of the difficulties that one must face in the general case. In addition, a quantitative higher dimensional version of this Theorem was key to [42]. We will use Calderón-Zygmund stopping data, to facilitate comparisons to the general case. This will also give an easier proof than is in [42, 58].

10.1. Off-Diagonal Estimates. We need a typical off-diagonal estimate, one that is far less refined than monotonicity principle.

Lemma 10.3. For all $0 < \alpha < \beta$, good intervals $J \subset I$, and function $f$ is supported off of $I$, there holds

\[(10.4) \quad |\langle T_{\alpha,\beta} \sigma f, g \rangle| \lesssim P(\sigma |f|, I)w(J)^{1/2} \|g\|_w.\]

for any function $g \in L^2(w)$, supported on $J$ and with integral zero.

Proof. Use the standard subtraction argument to see that

\[|\langle T_{\alpha,\beta} \sigma f, g(x) \rangle| = \left| \int \int_{R \setminus I} (K_{\alpha,\beta}(x,y) - K_{\alpha,\beta}(x_J,y))f(y)g(x) \sigma(dy)w(dx) \right| \leq \int \int_{J \setminus I} |x - x_J| \cdot |f(y)g(x)| \sigma(dy)w(dx).\]

The bound follows by Cauchy–Schwarz and inspection. \qed

10.2. The Global To Local Reduction. One need only prove that

\[|\langle T_{\sigma} P_{\text{good}}^\sigma f, P_{\text{good}}^w g \rangle| \leq \mathcal{J} \|f\|_\sigma \|g\|_w,\]

where $\mathcal{J} := \mathcal{T}_T + A_2^{1/2} + \mathcal{P}$. The set up is much like §4. (A) We will understand that one of our smooth truncations $T_{\alpha,\beta}(\sigma f)$ is actually being used, but systematically suppress it in the notation. (B) The $f$ and $g$ can be assumed to be good functions, (C) In fact, $f$ has the 'thin' Haar expansion in (4.1), and similarly for $g$, in order to reduce some case analysis below.

In analogy to (4.10), define

\[B_{\text{above}}(f, g) := \sum_{I: I \subset J} \sum_{J: J \subset I} E_{I}^\sigma I_{J} \Delta^w g \langle T_{\sigma} I_{J}, \Delta^w g \rangle_w,\]

and define $B_{\text{below}}(f, g)$ similarly. Since Lemma 4.3 depends only on the $A_2$ assumption, we have

Lemma 10.5. There holds

\[|\langle T_{\sigma} f, g \rangle_w - B_{\text{above}}(f, g) - B_{\text{below}}(f, g) | \leq A_2^{1/2} \|f\|_\sigma \|g\|_w.\]

Thus, the main technical result is

Lemma 10.6. There holds

\[|B_{\text{above}}(f, g)| \leq \mathcal{J} \|f\|_\sigma \|g\|_w.\]

The same inequality holds for $B_{\text{below}}(f, g)$.

In analogy to Definition 4.5, we define
**Definition 10.7.** Given any interval $I_0$, define $F_{\text{pivotal}}(I_0)$ to be the maximal subintervals $I \subseteq I_0$ such that

\begin{equation}
(10.8) \quad P[\sigma \cdot I_0, I]^2 \mathbb{w}(I) > 10^2 \sigma(I) .
\end{equation}

There holds $\sigma(\cup \{ F : F \in F(I_0) \}) \leq \frac{1}{4} \sigma(I_0)$, by the pivotal inequality \((10.1)\).

We make the following construction for an $f \in L^2_0(I_0, \sigma)$, the subspace of $L^2(I_0, \sigma)$ of functions of mean zero. Add $I_0$ to $F$, and set $\alpha_I(0) := 2^N |f|$. In the inductive stage, if $F \in F$ is minimal, add to $F$ those maximal descendants $F' \in D_F$, $F$ such that either (a) $B_F^w |f| \geq 10 \alpha_I(F)$, or (b) $F' \in F_{\text{pivotal}}(F)$. (Recall that $D_F$ is defined at \((4.1)\).) Then define

$$
\alpha_I(F') := \begin{cases} 
\alpha_I(F) & B_F^w |f| < 2 \alpha_I(F) \\
|f| & \text{otherwise}
\end{cases}
$$

We continue to use the notations $P_F^w f$ and $Q_F^w g$. Observe that Lemma 4.6 continues to hold for this choice of Calderón-Zygmund stopping data. And, in particular, the quasi-orthogonality condition \((4.9)\) holds.

In analogy to Corollary 4.11, there holds

**Lemma 10.9.** [The Global to Local Reduction] There holds

$$
|B_{\mathcal{F}, \mathcal{G}}^{\text{above}}(f, g)| \leq \mathcal{J} \| f \| \| g \|_w ,
$$

where $B_{\mathcal{F}, \mathcal{G}}^{\text{above}}(f, g) := \sum_{I, J : J \subset I} \mathbb{E}^g_{I, J} \Delta_F^g f \cdot (T_{g I}, \Delta^g f)'_w$.

**Proof.** The distinction between $J \subset I$ and $J \in I$ leads to these three cases.

A: $J \not\in \pi_F I \subset I$. ($J$ is strongly contained its $F$-parent.)

B: $J \not\in \pi_F I \not\in \pi_F J \subset I$. ($J$ is only strongly contained in it's $F$-grandparent, which is contained in $I$.)

C: $I \in J$, $J \not\in \pi_F I \subset I \subset \pi_F J$. ($J$ is not strongly contained in its $F$-parent, and $I$ is contained in the $F$-grandparent.)

Here we are using the notation that immediately proceeds Definition 7.1.

**Case A.** For integers $1 \leq s + t = u$, where $s = \lfloor u/2 \rfloor$, we have three intervals $F'' \subset F' \subset F$, where $J \in \pi_F I = F''$, and $\pi^2 F' = F$. We prove the bound below, valid for all $F \in \mathcal{F}$ and $u \in \mathbb{N}$:

\begin{equation}
(10.10) \quad \left| \sum_{F'' \in \mathcal{F} : \pi^2 F'' = F} B_{\mathcal{F}, \mathcal{G}}^{\text{above}}(P_F^w f, Q_{F''}^w g) \right| \leq 2^{-u/4} \mathcal{J} \cdot A \cdot B ,
\end{equation}

where $A^2 := \alpha_I(F)^2 \sigma(F) + \sum_{F' \in \mathcal{F} : \pi^2 F' = F} \alpha_I(F')^2 \sigma(F')$, and $B^2 := \sum_{F' \in \mathcal{F} : \pi^2 F' = F} \| Q_{F'}^w g \|_w^2$.

Then, quasi-orthogonality completes the bound in this case. (Of course this exponential gain along the stopping tree will not hold in absence of the pivotal condition.)

This variant of the 'Hilbert-Poisson exchange' argument is needed. We concentrate on intervals $I$ with $\pi_F I = F$, and $J$ with both $\pi_F J = F$ and $\pi_F I_J = F$. This is not assured, and we return to it at the end of
the proof. In the expression \( \mathbb{B}_I^g \Delta_I^g f \cdot \langle T_\sigma, \Delta_I^w g \rangle_w \), the argument of \( T_\sigma \) is written as \( I_I = (I_F - F') + F' \).

With \( F' \) as the argument of \( T_\sigma \), define real number \( \epsilon_I \) by

\[
\epsilon_I \alpha_\sigma(F) := \sum_{1: I_I \subseteq I} \mathbb{B}_I^g \Delta_I^g f.
\]

With the condition that the \( F \)-parent of \( I_I \) is \( F \), it follows that \( |\epsilon_I| \leq 1 \). And, we can write, for any interval \( F' \) with \( \pi^I_F F' = F \),

\[
\Phi(F') := \left| \sum_{1: \pi_F I_I \subseteq I} \sum_{\pi^I_F F' = F'} \sum_{j \in F''} \mathbb{B}_I^g \Delta_I^g f \cdot \langle T_\sigma F', \Delta_I^w g \rangle_w \right|
\]

\[
= \left| \langle T_\sigma F', \sum_{\pi^I_F F' = F'} \epsilon_I \Delta_I^w g \rangle_w \right|
\]

\[
\leq \mathcal{T} \alpha_\sigma(F) \sigma(F')^{1/2} \left[ \sum_{\pi^I_F F' = F'} \| Q_{F''}^w g \|_w^2 \right]^{1/2}.
\]

This depends upon the testing assumption on \( T_\sigma \) applied to intervals.

Then, observe that

\[
\sum_{\pi^I_F F' = F'} \sigma(F') \leq 5^{-s} \sigma(F) \leq 2^{-u + 1} \sigma(F),
\]

as follows from the construction of \( \mathcal{F} \). Therefore, by Cauchy–Schwarz, there holds

\[
\sum_{\pi^I_F F' = F'} \Phi(F') \leq 2^{-u/4} \mathcal{T} \cdot \alpha_\sigma(F) \sigma(F)^{1/2} \left[ \sum_{\pi^I_F F' = F'} \| Q_{F''}^w g \|_w^2 \right]^{1/2}.
\]

(This part of the argument will work in the general case, but there is no reason that the complementary sum will also have geometric decay.)

The complementary sum depends upon the estimate: For any \( F' \) with \( \pi^F F' = F \),

\[
\left| 1: \pi_F I_I \subseteq I \right| \mathbb{B}_I^g \Delta_I^g f \cdot \langle I_F - F' \rangle(x) \right| \leq \alpha_\sigma(F) \cdot F(x).
\]

Then, the Hilbert–Poisson exchange argument continues. Apply (10.4) to see that

\[
\Psi(F') := \left| \sum_{1: \pi_F I_I \subseteq I} \sum_{\pi^I_F F' = F'} \sum_{j \in F''} \mathbb{B}_I^g \Delta_I^g f \cdot \langle T_\sigma(I_I - F''), \Delta_I^w g \rangle_w \right|
\]

\[
\leq \alpha_\sigma(F) \sum_{\pi^I_F F' = F'} \mathcal{P}(\sigma(F - F'), F'') w(F'')^{1/2} \| Q_{F''}^w g \|_w
\]

In the estimate below, we use Cauchy–Schwarz, and critically, the estimate of Lemma 6.1, with integer s, giving a geometric decay below.
\[ \leq 2^{-(1-2\varepsilon)s} \alpha_f(F)P(\sigma \cdot F, F')w(F')^{1/2}\left[ \sum_{F'' \in \mathcal{F}} \|Q_{F''}w\|^2\right]^{1/2} \]

Another application of Cauchy–Schwarz and an appeal to the pivotal condition (10.1) will show that

\[ \sum_{F' \in \mathcal{F}} \Psi(F') \leq 2^{-(1-2\varepsilon)u/2} \alpha_f(F)\sigma(F)^{1/2}\left[ \sum_{J: \pi_{F'}=F} \|Q_{J}w\|^2\right]^{1/2}. \]

For \(0 < \varepsilon < \frac{1}{4}\), which we have assumed throughout, we have desired estimate. This completes the proof of (10.10), with however the additional restriction that we summed over pairs \((I, J)\) in Case A, with \(\pi_{F'}I = \pi_{F'I}\).

It remains to prove the estimate below

\[ \left| \sum_{F' \in \mathcal{F}} \sum_{J: \pi_{F'}=F} \sum_{\tilde{F}' \in \mathcal{F}} \mathbb{E}_{I,J,\tilde{F}'} \Delta_{\tilde{F}'} f \cdot \langle T_{\tilde{F}} \Delta_{\tilde{F}'} w \rangle \right| \leq 2^{-u/4}J \left( \sum_{F' \in \mathcal{F}} \alpha_f(F')^2\sigma(F') \right)^{1/2} \|g\|_w. \]

But this only depends upon stopping, and interval testing, so we omit the details.

**Case B.** Note that if \(I \notin \pi_{F}J\), then since \(\mathcal{F}\) is a subtree of \(D_\varepsilon\), as defined in (4.1), it is then necessary that \(\pi_{F'}J \notin \pi_{F'}^2J\). Thus, we see that the argument of Case A can then be easily adapted to this case. We omit the details.

**Case C.** In this case, we fix \(F \in \mathcal{F}\), and consider \(F' \) with \(\pi_{F'}F' = F\). The first stage of the Hilbert-Poisson exchange argument is as follows. Again we have the distinction arising from \(I\) and the child of \(I\) have different parents. Below, we consider the same parents. Below, the argument of \(T_{\sigma}\) is \(F\).

\[ \left| \sum_{I: \pi_{F'}=F} \sum_{I: \pi_{F'}=F} \sum_{\tilde{F}' \in \mathcal{F}} \mathbb{E}_{I,J,\tilde{F}'} \Delta_{\tilde{F}'} f \cdot \langle T_{\tilde{F}} \Delta_{\tilde{F}'} w \rangle \right| \]

\[ = \alpha_f(F)\langle T_{\sigma}(F), \sum_{F': \pi_{F'}=F} \|Q_{J}w\|^2\rangle \]

\[ \leq \mathcal{J} \alpha_f(F)\sigma(F)^{1/2}\left[ \sum_{F': \pi_{F'}=F} \|Q_{F''}w\|^2\right]^{1/2}. \]

Here, the constants \(\varepsilon_{J}\) are bounded by an absolute constant, by an argument that uses the stopping values to bound telescoping sums of martingale differences. The quasi-orthogonality argument (4.9) bounds the sum over \(F\) of this expression.

The complementary sum is formed as above, but the argument of \(T_{\sigma}\) is now \(F - I_{\tilde{J}}\). We will further hold the lengths of \(J\) relative to it’s parent \(F'\) fixed; it can only take one of \(r\) values since \(J \notin F'\). Note
that for $F'$ as above, and $\pi_F J = F'$ and $J \not\subset F'$, we have
\[
\left| \sum_{1: \pi_F I = F} \sum_{J: \pi_F J = F'} P^g_{I, J} \Delta^0 f \cdot (T_\sigma F - I_J, h_J^w) \right| \leq \alpha_f(F) \cdot F(x).
\]

Hence, by (10.4), we can estimate below, where we write $r = s + t$, for two non-negative integers $s, t$.

For each $F' \in \mathcal{F}$ with $\pi_F F' = F$,
\[
\left| \sum_{1: \pi_F I = F} \sum_{J: \pi_F J = F'} \sum_{I \supset F', 2^{|F'|} \leq |I| 2^{|J|} = |F|} P^g_{I, J} \Delta^0 f \cdot (T_\sigma F - I_J, h_J^w) \right| \leq \alpha_f(F) \sum_{J: \pi_F J = F'} P(\sigma \cdot F, J) w(J)^{1/2} |\hat{g}(J)|.
\]

Sum this over $F'$, use Cauchy–Schwarz and the pivotal property to see that it at most
\[
\alpha_f(F) \sigma(F)^{1/2} \left( \sum_{F': \pi_F F' = F} \varepsilon_{F'} \|Q_{F'} g\|_w^2 \right)^{1/2}.
\]

Then, appeal to quasi-orthogonality to finish this last case, and hence the proof of the global to local reduction.

\[\square\]

10.3. The Local Estimate. It remains to prove the following local estimate:
\[
|B_{F}^{\text{above}}(P_{F}^g f, g)| \leq \mathcal{T} \{ \alpha_f(F) \sigma(F)^{1/2} + \|P_{F}^g f\|_{\sigma} \|g\|_w \}, \quad Q_{F}^w g = g,
\]

for then quasi-orthogonality will complete the bound on $B_{F}^{\text{above}}(f, g)$.

In the bilinear form above, the argument of $T_\sigma$ is, for a pair of intervals $J \subset I$, $I_J = (F - I_J) + F$. Using linearity, and focusing on the argument of $T_\sigma$ being $F$, we can repeat the argument of (4.19), which depends upon the fact that the averages of $f$ are controlled. Below, there is an requirement that $I_J$ has $F$-parent $F$, which we are free to add since $Q_{F}^w g = g$.

This bound follows the argument of (4.19), and we suppress the details.

It therefore remains to consider the stopping form
\[
B_{F}^{\text{stop}}(f, g) := \sum_{1: \pi_F I = F} \sum_{J: \pi_F J = F} P^g_{I, J} \Delta^0 f \cdot (T_\sigma (I_0 - I_J), \Delta^0 g)w.
\]

Lemma 10.11. For all $F \in \mathcal{F}$, there holds
\[
|B_{F}^{\text{stop}}(f, g)| \leq D \|f\|_{\sigma} \|g\|_w.
\]

Proof. This depends very much on the selection of stopping intervals. In fact there is geometric decay, holding the relative lengths of $I$ and $J$ fixed. Estimate for integers $s \geq r$,
\[
\sum_{1: \pi_F I = F} \sum_{J: \pi_F J = F} \sum_{|I| = 2^s |J|} |P^g_{I, J} \Delta^0 f \cdot (T_\sigma (I_0 - I_J), \Delta^0 g)w| \leq \sum_{1: \pi_F I = F} \sum_{\theta \in \{\pm\}} \sum_{|I| = 2^s |J|} |\hat{f}(I)| |\hat{\sigma}(I_0)|^{1/2} \sum_{J: \pi_F J = F} P(\sigma(F - I_0), J) |\langle \hat{g}, h_J^w \rangle| |\hat{g}(J)|.
\]
\[
\sum_{f \in F} \left[ \sum_{I : \pi F \cap I = F} \hat{f}(I) \right]^{1/2} \times \left[ \sum_{J : |J| \geq 2^s |J|} \frac{1}{\sigma(I)} P(\sigma(F - I_\theta), J)^2 w(J) \right].
\]

Here, we have used (a) used the bound \( |\sum G \pi I_\theta \Delta f| \leq \sum |\hat{f}(I)|/|I_{\theta}|^{1/2} \); (b) appealed to (10.4); (c) used Cauchy–Schwarz, together with the fact that for \( J \in F \), there is a unique \( I_\theta \) containing it, with length \( 2^s |J| \).

It remains to bound \( M_s \), gaining a geometric decay in \( s \), and appealing to the pivotal condition. Return to the inequality (6.2), to gain the geometric decay,

\[
\sum_{J : |J| = 2^s |J|} P(\sigma(F - I_\theta), J)^2 w(J) \leq 2^{-(1-\varepsilon)s} P(\sigma \cdot F, I_\theta)^2 w(I_\theta) \leq 2^{-(1-\varepsilon)s} 2^s w(I_\theta),
\]

where the decisive point is that \( I_\theta \) has \( F \)-parent \( F \), hence it must fail the inequality (10.8). \( \Box \)

11. Example Weights

The sharpness of the different conditions in the main theorem is the subject of this section.

**Theorem 11.1.** There are pairs of weights \( \sigma, w \), with no common point masses, that satisfy any one of these conditions.

1. The pair of weights satisfies the full Poisson \( A_2 \) condition, but the norm inequality for the Hilbert transform (1.1) does not hold.
2. The pair of weights satisfies the full Poisson \( A_2 \) condition, and the testing inequality (1.5), but the norm inequality for the Hilbert transform (1.1) does not hold.
3. The pair of weights satisfy the two weight norm inequality (1.1), but not the pivotal condition (3.19).

Point (1) is a counterexample to Sarason’s Conjecture, first disproved by Nazarov [31]. In contrast to his argument, an explicit pair of weights are exhibited.

11.1. The Initial Steps in the Main Construction. Let \( C = \bigcap_{n=0}^{\infty} C_n \) be the standard middle third Cantor set in the unit interval. Thus, \( C_0 = [0, 1], C_1 = [0, 1/3] \cup [2/3, 1], \) and more generally

\[
C_n = \bigcup \left\{ [x, x + 3^{-n}] : x = \sum_{j=1}^{n} \varepsilon_j 3^{-j}, \varepsilon_j \in \{0, 2\} \right\}.
\]

Let \( w \) be the standard uniform measure on \( C \). Thus \( w(I) = 2^{-n} \) on each component of \( C_n, n \in \mathbb{N}_0 \). This is phrased slightly differently. Let \( \mathcal{K} \) be the collection of components of all the sets \( C_n \). Then, for each \( K \in \mathcal{K} \), there holds \( w(K) = |K|^{1/3} \).

The weight \( \sigma \) will be a sum of point masses selected from the intervals in \( \mathcal{G} \), taken to be the components of the open set \( [0, 1] - C \). \((G \) is for 'gap.' \) Consider the \( Hw \) restricted an interval \( G \in \mathcal{G} \). This is a smooth, monotone function, hence it has a unique zero \( z_G \). Then, the weight \( \sigma \) is

\[
\sigma := \sum_{G \in \mathcal{G}} s_G \cdot \delta_{z_G},
\]
Figure 10. The approximates to the Cantor set $C$ on the left, and on the right, the gaps, namely the components of $[0,1] - C$. The intervals on the left are in $\mathcal{K}$, and those on the right are in $\mathcal{G}$.

Figure 11. The selection of the points $z_G$ and $z'_G$ for a gap interval $G$. The function $H_w$, restricted to $G$ is monotone increasing, hence has a unique zero, the point $z_G$. The second point $z'_G$ will be to the right, but a distance to the boundary of $G$ that is at least $c|G|$, for absolute constant $0 < c < \frac{1}{2}$.

where $s_G > 0$ will be chosen momentarily, consistent with the $A_2$ condition. A second measure is given by $\sigma' := \sum_{G \in \mathcal{G}} s_G \cdot \delta_{z'_G}$, where $z'_G$ is the unique point in $G$ at which $H_w(z'_G) = |G|^{-1 + \frac{\ln 2}{\ln 3}}$. See Figure 11.

The constants $s_G$ are be specified by the simple $A_2$ ratio

$$\frac{w(3G)}{|G|} \cdot \frac{\sigma(G)}{|G|} = 2,$$

that is $s_G = 2|G|^{2 - \frac{\ln 2}{\ln 3}}$.

To see this, note that

$$w(3G) = w(G - |G|) + w(G + |G|) = 2|G|^{\frac{\ln 2}{\ln 3}},$$

since $G \pm |G|$ are components of some $C_n$. With this definition, the basic facts about the $w$ and $\sigma$ come from the geometry of the Cantor set and the the relations below,

\begin{equation}
(11.2) \quad \begin{aligned}
w(I) & \leq |I|^{\frac{\ln 2}{\ln 3}}, & \text{I is triadic}, \\
\sigma(I) & \leq |I|^{2 - \frac{\ln 2}{\ln 3}}, & \text{I is triadic, I not strictly contained in any } G \in \mathcal{G}.
\end{aligned}
\end{equation}

On the other hand, if $I \in \mathcal{G} \cup \mathcal{K}$, the inequalities above can be reversed, namely

\begin{equation}
(11.3) \quad \begin{aligned}
w(3I) & \simeq |I|^{\frac{\ln 2}{\ln 3}}, & \text{I is triadic,} \\
\sigma(I) & \simeq |I|^{2 - \frac{\ln 2}{\ln 3}}, & I \in \mathcal{G} \cup \mathcal{K}.
\end{aligned}
\end{equation}

The properties of these measures that we are establishing are as follows.
Lemma 11.4. For the measures just defined, there holds

1. The Hilbert transform $H_\sigma$ is bounded from $L^2(\sigma)$ to $L^2(w)$.

2. The Hilbert transform $H_\sigma$ is unbounded from $L^2(\sigma')$ to $L^2(w)$, but the pair of weights satisfy the $A_2$ condition, and the testing conditions

$$\sup_{I \text{ an interval}} \sigma'(I)^{-1} \int_I |H_\sigma'I|^2 \, dw < \infty.$$  

Concerning point 2, the unboundedness of $H_w$ is direct from the construction of $\sigma'$.

$$\int (H_w)^2 \, d\sigma' = \sum_{G \in \mathcal{G}} H_w(z'_G)^2 \sigma'([z'_G])$$

(11.5)

$$= \sum_{G \in \mathcal{G}} |G|^{2-\frac{\ln 2}{\ln 3} - 2(1-\frac{\ln 2}{\ln 3})} = \sum_{G \in \mathcal{G}} |G|^{|\ln 2|} \ln 3 = \infty.$$  

There are exactly $2^{n-1}$ elements of $\mathcal{G}$ of length $3^{-n}$, proving the sum is infinite.

11.2. The Poisson $A_2$ Condition.

Lemma 11.6. For either weight $\mu \in \{\sigma, \sigma'\}$, the pair of weights $w, \mu$ satisfy the $A_2$ condition.

Proof. It suffices to check the $A_2$ condition on the the triadic intervals in the unit interval. Let us begin by showing that for any triadic interval $I \in \mathcal{K} \cup \mathcal{G}$,

(11.7) $P(\sigma, I) \leq \frac{\sigma(I)}{|I|}$, and $P(w, I) \leq \frac{w(3I)}{|I|}$.

For then, the control of the simple $A_2$ ratio will imply the control of the full $A_2$ ratio. (For the inequality on $w$, the triple of the interval appears on the right, since $w(I)$ can be zero if $I \in \mathcal{G}$.) Now, it will be clear that this argument is insensitive to the location of the points $z_G$ and $z'_G$, so the same argument for $\sigma$ will work equally well for $\sigma'$.

Let us consider $\sigma$. Using (11.3), there holds

$$P(\sigma, I) \leq \frac{\sigma(I)}{|I|} + \sum_{k=1}^{\infty} \int_{3^kI \setminus 3^{k-1}I} \frac{|I|}{|I|^2 + \text{dist}(x, I))^2} \sigma(dx)$$

$$\leq \frac{\sigma(I)}{|I|} + \sum_{k=1}^{\infty} \sigma(3^kI)$$

$$\leq \frac{\sigma(I)}{|I|} + \sum_{k=1}^{\infty} 3^{-k} |3^kI|^{1-\frac{\ln 2}{\ln 3}} \leq \frac{\sigma(I)}{|I|} \sum_{k=0}^{\infty} 3^{-k} \ln 3 \leq \frac{\sigma(I)}{|I|}.$$  

Turning to the weight $w$, one has

$$P(w, I) \leq \frac{w(3I)}{|I|} + \sum_{k=2}^{\infty} \int_{3^kI \setminus 3^{k-1}I} \frac{|I|}{|I|^2 + \text{dist}(x, I))^2} \sigma(dx)$$

$$\leq \frac{w(3I)}{|I|} + \sum_{k=2}^{\infty} \frac{w(3^kI)}{3^k |3^kI|}.$$
\[ \sum_{k=2}^{\infty} 3^{-k} |3^k I|^{-1 + \frac{\ln 2}{\ln 3}} \leq \frac{w(3I)}{|I|} \sum_{k=1}^{\infty} 3^{-k \left( \frac{\ln 2}{\ln 3} \right)} \leq \frac{w(3I)}{|I|}. \]

The $A_2$ product $P(\sigma, I) \cdot P(w, I)$ has been bounded for $I \in K \cup G$. Suppose that $I$ is a triadic interval that is not in these two collections. Then, $I$ must be strictly contained in some gap $G \in G$. Writing $I^{(k)} = G$, where, $I^{(k)}$ denotes the $k$-fold parent of $I$ in the triadic grid, we have $w(G) = 0$. Hence,

\[ P(w, I) = \int_{[0,1] \setminus G} \frac{|I|}{(|I| + \text{dist}(x, I))^2} w(dx) \simeq 3^{-k} P(w, G). \]

First, consider $\sigma$ restricted to the gap $G$:

\[ P(w, I)P(\sigma \cdot G, I) \leq 3^{-k} P(w, G) \frac{\sigma(G)}{|I|} \simeq P(w, G) \frac{\sigma(G)}{|G|} \leq 1. \]

Now, we have to consider the Poisson average of $\sigma$ off of the gap $G$, in which case we have

\[ P(\sigma \cdot ([0,1] \setminus G), I) \simeq 3^{-k} P(\sigma, G), \]

and so the estimate follows.

\[ \square \]

### 11.3. The Testing Conditions

We turn to the testing conditions, using in an essential way the precise definition of the weight $\sigma$: it gives a huge cancellation, which simplifies things considerably.

**Lemma 11.8.** For any interval $I$, there holds

\[ \int_I |H_w I|^2 \, d\sigma \lesssim w(I). \]

**Proof.** By construction of $\sigma$, there are two reductions. The first is simple, namely that the two endpoints of the interval $I$ can be taken to be an endpoint of an interval in $G$. The second comes from the construction of $\sigma$: $Hw \equiv 0$, relative to $d\sigma$ measure. Hence,

\[ \int_I |H_w I|^2 \, d\sigma = \int_I H_w([0,1] - I)^2 \, d\sigma, \]

namely the complement of $I$ is the argument of the Hilbert transform on the right.

Then, one abandons all further cancellations. Let us show that for all intervals $K \in K$ (the components of the sets $C_n$ which generate the Cantor set),

\[ \int_K |H_w K_{rt}|^2 \, d\sigma \lesssim w(K), \tag{11.9} \]

where $K_{rt}$ is the right component of $[0,1] \setminus K$. The same estimate holds for the left component, and this completes the proof. For, if we set $I_{rt}$ to be the right component of $[0,1] \setminus I$, and take $K^1, K^2, \ldots$, to be the maximal intervals in $K$ contained in $I$, there holds

\[ \int_I (H_w I_{rt})^2 \, d\sigma \leq \sum_{n=1}^{\infty} \int_{K^n} (H_w K^n_{rt})^2 \, d\sigma \leq \sum_{n=1}^{\infty} w(K^n) \lesssim w(I). \]
Now, for $K \in \mathcal{K}$, let $K_1, K_2, \ldots$, be the maximal intervals in $\mathcal{K}$ that lie to the right of $K$. Arranging them in increasing length, note that the length of $K_1$ is either $|K|$ or $3|K|$. For $n \geq 2$, the length of $K_n$ increases by a factor of $3$, and $\text{dist}(K, K_n) \gtrsim |K_n|$, and hence there are at most $1 - \log_3|K|$ such intervals in $\mathcal{K}$. Here is an illustration:

| $K$ | $K_1$ | $K_2$ | ... | $K_3$ |

Then, one has the estimate below, where the sum is of a decreasing geometric series, estimated by its first term.

$$|Hw_{Kr}| \lesssim \sum_{n=1}^{\infty} \frac{w(K_n)}{|K_n|} \approx \frac{w(K)}{|K|}.$$ 

Hence, (11.9) follows from the control of the $A_2$ ratio.

An important part of the remaining arguments is that points $z_G$ and $z_G'$ cannot cluster close to the boundary of $G$.

**Lemma 11.10.** There is a constant $0 < c < \frac{1}{2}$ such that

$$|z_G - z_G'| \leq c|G|.$$ 

**Proof.** Estimate $Hw$ at the midpoint $z_G''$ of a component $G$. By symmetry of the Hilbert transform, and the Cantor set, it always holds that $H(w1_{3G})(z_G') = 0$, so that appealing to (11.2),

$$|Hw(z_G'')| = |H(w1_{3G})(z_G'')| \leq \sum_{k=2}^{n} \frac{w(3^kG)}{|3^kG|} \lesssim \sum_{k=2}^{n} |3^kG|^{-1 + \frac{\ln 2}{\ln 3}} \lesssim |G|^{-1 + \frac{\ln 2}{\ln 3}}.$$

Next, we turn to a derivative calculation. The function $Hw$, restricted to $G$ is a smooth function, one that diverges at the end points of $G$ at a rate that reflect the fractal dimension of $G$. For any $x \in G$, note that

$$\frac{d}{dx}Hw(x) = \int_{\mathbb{C}} \frac{w(dy)}{(y-x)^2} \gtrsim \frac{w(3G)}{|G|^2} \simeq |G|^{-2 + \frac{\ln 2}{\ln 3}}.$$ 

This is a uniform lower bound, and in fact the lower bound is very poor at the boundaries of $G$. Indeed,

$$\frac{d}{dx}Hw(x) \gtrsim \text{dist}(x, \partial G)^{-2 + \frac{\ln 2}{\ln 3}}.$$ 

It follows that we have $|z_G - z_G'| < c|G|$, for some $0 < c < \frac{1}{2}$. That is, one need only move at fixed small multiple of $|G|$, passing from the location of the zero $z_G$ to the point $z_G'$.

The second half of the testing intervals inequalities is as follows.
Lemma 11.11. For $\mu \in \{\sigma, \sigma'\}$, and any interval $I$,

$$\int_I |H_{\mu} I|^2 \, dw \lesssim \mu(I). \tag{11.12}$$

Proof. For the sake of specificity, let $\mu = \sigma$. Indeed, by Lemma 11.10, the same argument will work for $\sigma'$. To fix ideas, let us assume that $I \in \mathcal{K}$. Write the left, middle and right thirds of $I$ as $I_{-1}, I_0, I_1$, respectively. Then, note that

$$\int_I H_{\sigma}(I)^2 \, dw = \int_{I_{-1} \cup I_1} H_{\sigma}(I)^2 \, dw \tag{11.13} \leq \int_{I_{-1} \cup I_1} H_{\sigma}(I_0)^2 \, dw + \int_{I_0} H_{\sigma}(I_0 + I_1)^2 \, dw + \int_{I_1} H_{\sigma}(I_{-1} + I_0)^2 \, dw + \int_{I_1} H_{\sigma}(I_1)^2 \, dw. \tag{11.14}$$

The first term on the right is simple. On the interval $I_0$, $\sigma$ is a point mass, at a point that is at distance $\geq c|I|$ from $I_{\pm 1}$. Thus, by (11.3),

$$\int_{I_{-1} \cup I_1} H_{\sigma}(I_0)^2 \, dw \lesssim \left| \frac{|I|^{4-2\ln \frac{2}{3}}}{|I|^2} \right| \left| \frac{|I|^{\ln \frac{2}{3}}}{|I|^2} \right| \simeq \sigma(I).$$

That completes the first integral. The remaining two integrals in (11.13) are handled by a similar argument.

Concerning the two integrals in (11.14), one should note that $I_{\pm 1} \in \mathcal{K}$ and that $\sigma(I_{\pm 1}) \leq 3^{-2+2\ln \frac{2}{3}} \sigma(I)$. This geometric factor is smaller than $\frac{1}{2}$, therefore one can recurse on (11.13) and (11.14) to see that

$$\int_{\mathcal{K}} H_{\sigma}(I)^2 \, dw \lesssim \sigma(I), \quad K \in \mathcal{K}. \tag{11.15}$$

For a general interval $I$, since $\sigma$ is a sum of Dirac masses, we can assume that the interval $I$ is in a canonical form. Namely, each endpoint of $I$ can be assumed to be an endpoint of an interval in $\mathcal{G}$. The basic inequality is

$$\sum_{K \in \mathcal{K}_I} \int_K |H_{\sigma}(I - K)|^2 \, dw \lesssim \sigma(I), \tag{11.16}$$

where $\mathcal{K}_I$ is the maximal elements of $\mathcal{K}$ contained in $I$. The integration is over $K$, and the argument of the Hilbert transform is $I - K$.

To see that (11.16) implies the Lemma, note that by (11.15),

$$\int_I H_{\sigma}(I)^2 \, dw = \sum_{K \in \mathcal{K}_I} \int_K H_{\sigma}(I)^2 \, dw \leq \sum_{K \in \mathcal{K}_I} \int_K H_{\sigma}(I - K)^2 \, dw + \sum_{K \in \mathcal{K}_I} \int_K H_{\sigma}(K)^2 \, dw \lesssim \sigma(I) + \sum_{K \in \mathcal{K}_I} \sigma(K) \lesssim \sigma(I).$$
In fact, (11.16) follows from
\begin{equation}
(11.17) \int_{K} |H_{\sigma}(I - K)|^{2} \, dw \leq \frac{\sigma(I)^{2}}{|I|^{2}} w(K), \quad K \in \mathcal{K}_{1}.
\end{equation}

For this is summed over $K \in \mathcal{K}_{1}$, and then one uses the $A_2$ property.

To prove (11.17), all hope of cancellation is abandoned. For an interval $K \in \mathcal{K}_{1}$, let us consider component $I_{rt}$ of $I - K$ which lies to the right of $K$. It has a Whitney like decomposition into a finite sequence of intervals $J_{1}, \ldots, J_{t}$ that we construct now. These intervals will have the property that they are (a) pairwise disjoint, (b) their union is $I_{rt}$, (c) and dist($K, \text{supp}(\sigma J_{s})$) $\geq |J_{s}| \geq \frac{3}{2} |K|$, for all $1 \leq s \leq t$.

Now, $J_{1} = K + |K| \in \mathcal{G}$. If this interval is not contained in $I$, it follows that $K$ contains the right hand endpoint of $I$, and there is nothing to prove. Assuming that $J_{1} \subset I$, the inductive step is this. Given $J_{1}, \ldots, J_{s}$, as above, whose union is not $I_{rt}$

1. If $J_{s} \in \mathcal{G}$, then $J_{s} + |J_{s}| \in \mathcal{K}$. If this interval is contained in $I_{rt}$, then we take $J_{s+1} = J_{s} + |J_{s}| \in \mathcal{K}$, and repeat the recursion. Otherwise, we update $J_{s} = I_{rt} - \bigcup_{u=1}^{s-1} J_{t}$, and the recursion stops.

2. If $J_{s} \in \mathcal{K}$, then it follows that $J_{s-1} \in \mathcal{G}$, and the element of $\mathcal{G}$ immediately to the right of $J_{s}$ is $3(|J_{s} + 6|J_{s}|)$. If this interval is contained in $I_{rt}$, then we take $J_{s+1} = 3(|J_{s} + 6|J_{s}|) \in \mathcal{G}$, and repeat the recursion. Otherwise, we update $J_{s} = I_{rt} - \bigcup_{u=1}^{s-1} J_{t}$, and the recursion stops.

With this construction, it follows that

\[ |H_{\sigma}(I_{rt}) \cdot K| \leq \sum_{u=1}^{t} \frac{\sigma(J_{s})}{|J_{s}|} \leq \sum_{n=1}^{\infty} |J_{s}|^{-\frac{\ln n}{\ln 3}} \leq \frac{\sigma(I)}{|I|}. \]

This proves the ‘right half’ of (11.17), that is, when the argument of the Hilbert transform is $I_{rt}$. The ‘left half’ is the same, so the proof is complete. \hfill \Box

At this point, we have proven that the pair of weights $(w, \sigma')$ satisfy the full Poisson $A_2$ condition, and the testing condition (11.12). But, $\|H w\|_{L^{2}(\sigma')}^{}$ is infinite, by (11.5). Hence, points (1) and (2) of Theorem 11.1 are shown.

We have also shown that the pair of weights $(w, \sigma)$ satisfy the full Poisson $A_2$ condition, and both sets of testing conditions. Hence, by our main theorem, $Hw$ is bounded from $L^{2}(w)$ to $L^{2}(\sigma)$. This pair of weights also fail the pivotal condition (3.19) of Nazarov-Treil-Volberg [35]. This is verified by observing that the collection $\mathcal{G}$ of gaps is a partition of $[0, 1]$, and

\[ \sum_{G \in \mathcal{G}} P(w, G)^{2}w(G) \simeq \sum_{G \in \mathcal{G}} \frac{w(3G)^{2}}{|G|^{2}} \sigma(I) \simeq \sum_{G \in \mathcal{G}} w(3G) \simeq \sum_{G \in \mathcal{G}} |G|^{-\ln 3} = \infty \]

since $\mathcal{G}$ contains $2^{n}$ intervals of length $3^{-n}$, for all integers $n$. Here, we have used (11.7), followed by (11.2). Since $\inf_{x \in G} Mw(x) \geq P(w, G)$, this also shows that the maximal function $M$ is not bounded from $L^{2}(w)$ to $L^{2}(\sigma)$.

Notice in contrast that the energy inequality (3.18) for the partition $\mathcal{G}$ is trivial, since $\sigma$ restricted to any interval $G$ is a point mass, hence $E(\sigma, G) = 0$, for all $G \in \mathcal{G}$.

11.4. Context and Discussion.
11.4.1. Counterexamples were an important source of inspiration on these questions. The early paper of Muckenhoupt and Wheeden [30] includes an example of the fact that the simple $A_2$ condition is not sufficient for the two weight inequality. For instance, the boundedness of the simple $A_2$ ratio is simple to check for the pair $w = \delta_0$, and $\sigma(dx) = x1_{[0,\infty)}dx$. Then, one sees that for $f = \frac{1}{x}1_{[1,L]}$, 
$$\sqrt{\log L} \lesssim \|f\|_\sigma \ll \log L \lesssim \|H_\sigma f\|_w, \quad L > 1.$$ 
Thus, the Hilbert transform is unbounded. And, one can directly see that the half-Poisson $A_2$ condition fails.

Much harder, is the fact that the Poisson $A_2$ condition is not sufficient. This was the contribution of Nazarov [31]. This example lead to the conjecture of Nazarov-Treil-Volberg [58] proved herein. A more delicate example, of a pair of weights which satisfied the Poisson $A_2$ condition, and one set of testing conditions, say (1.5), but not the norm inequality was that of Nazarov-Volberg [37]. Also see Nikolski-Treil [39], for a related example in to disprove a conjecture about similarity to a normal operator, as shown by Nikolski-Treil [39]. Both of these latter examples were based upon Nazarov’s indirect example.

11.4.2. The example given here is directly inspired by a Cantor set type example in Sawyer’s two weight maximal function paper [49]. It is drawn from [20], with the purpose to show that the pivotal condition of Nazarov-Treil-Volberg [35, 58] was not necessary for the two weight inequality to hold. This was an explicit example, and also pointed to the primary role of the notion of energy. It is very interesting and delicate, in that the point masses have to be placed on the zeros of the Hilbert transform, in order to obtain the boundedness of the transform. It is also humbling in that it still does not reveal how delicate the proof of the sufficiency in the main theorem needs to be.

11.4.3. It is subtle example of Maria Carmen Reguera [44] and Reguera-Thiele [46] that proves this, as is pointed out by Reguera-Scurry [45].

**Theorem 1.** There is a pair of weights for which the maximal function $M_\sigma$ is bounded from $L^2(\sigma) \rightarrow L^2(w)$ and $M_w$ is bounded from $L^2(w) \rightarrow L^2(\sigma)$, but norm inequality for the Hilbert transform (1.1) does not hold.

This is quite a bit more intricate than the examples we have presented. It had been suggested, in the early days of the weighted theory, that the boundedness of the maximal functions would be sufficient for the norm boundedness of the Hilbert transform. On the other hand, if one considers ‘off-diagonal’ estimates, then boundedness of the maximal function is sufficient for norm inequalities for singular integrals [8].

12. Applications of the Main Inequality

The interest in the two weight problem stems from a range of potential applications arising in sophisticated arenas of complex function and spectral theory. The motivations for these questions are complicated, and based upon subtle theories. The connections to the two weight Hilbert transform are not always immediate, and the properties of interest are frequently more intricate than those of mere boundedness of a transform. Nevertheless, the acknowledged experts Belov-Mengestie-Seip in [3] write “...we have found it both useful and conceptually appealing to transform the subject into a study of the mapping properties of discrete Hilbert transforms. We have learned to appreciate that the essential difficulties thus seem to appear in a more succinct form.” A brief guide to the subjects, and some of the ‘essential difficulties’ follow.
12.1. **Sarason’s Question on Toeplitz Operators.** This question arose from Sarason’s work on exposed points of $H^1$ [47]. Indeed, this was part of an influential body of work that pointed to the distinguished role of de Branges spaces in the subject. This paper contains examples of pairs of functions $f, g$, for which the individual Toeplitz operators were unbounded, but the composition bounded.

**Question 12.1 (Sarason [48]).** Characterize those pairs of outer functions $g, h \in H^2$ for which the composition of Toeplitz operators $T_g T_h$ is bounded on $H^2$.

Following [48], for a function $h \in L^2(T)$, the Toeplitz operator $T_h$ can be thought of as taking $f \in H^2$ to the space of analytic functions by the definition

$$T_h f(z) := \frac{1}{2\pi} \int_{\partial D} f(e^{i\theta})h(e^{i\theta})k_z(e^{i\theta}) \, d\theta,$$

where $k_z(w) := \frac{(1-|w|^2)^{1/2}}{1-wz}$ is the reproducing kernel.

Also in [48] is an argument of S. Treil that a Poisson $A_2$ condition is necessary condition for the boundedness of the composition:

$$\sup_{z \in \mathbb{D}} |Pf|^2(z) |Pg|^2(z) < \infty,$$

where $P$ denotes the Poisson extension to the unit disk. Sarason wrote that ‘It is tempting to conjecture that the last condition is also sufficient for the boundedness of $T_g T_h$.’ This statement, widely referred to as the Sarason Conjecture, is of interest in both the Hardy and Bergman space settings. (In the latter case, see the striking resolution, due to Aleman-Pott-Reguera [2].)

The connection with the two weight problem for the Hilbert transform is indicated by the diagram from [7, §5], see Figure 12. In the diagram, $M_h$ is multiplication by $h$ and $P_+$ is the Riesz projection from $L^2$ to $H^2$. The boundedness is equivalent to

$$M_g P_+ M_f : H^2 \mapsto H^2.$$

The structure of outer functions leads to these simplifications. Since the product of analytic is analytic, the second $H^2$ above can be replaced by $L^2$, and then, the outside multiplication $M_g$ can then be replaced by $M_{|g|}$. Thus, we are considering $M_{|g|} P_+ M_{|f|} : H^2 \mapsto L^2$. Now, $\overline{T}$ is anti-analytic, so we can replace $H^2$ above by $L^2$. Moreover, the multiplication operator $M_{f/|f|}$ is unitary, since an outer function can be equal to zero on $T$ only on a set of measure zero. Thus, it is equivalent to consider

$$M_{|f|} P_+ M_{|f|} : L^2 \mapsto L^2.$$

This is a two weight inequality for $P_+$.\(^7\)

The Riesz projection is a linear combination of the identity and the Hilbert transform, and our main theorem will apply to it. Note that the inequality

$$\|P_+(|f|\phi)\|_{L^2(|g|^2 dx)} \leq \|\phi\|_{L^2(dx)}$$

is equivalent to

$$\|P_+(|f|^2\psi)\|_{L^2(|g|^2 dx)} \leq \|\psi\|_{L^2(|f|^2 dx)} \cdot$$

---

\(^7\)Sergei Treil helped us with the history of this question.
Recall that $P_+ = I - \frac{1}{\pi} H$, according to how we defined the Hilbert transform, where $I$ represents the identity operator. In the two weight setting, we interpret the norm inequality $\|P_+ (\sigma f)\|_w \leq \|f\|_\sigma$, as uniform over all truncations $0 < \tau < 1$ defined by

$$P_{+\tau}(\sigma f) := \sigma f + \frac{i}{\pi} \int_{\tau < |x-y| < \tau^{-1}} f(y) \frac{\sigma(dy)}{y-x}$$

**Theorem 12.2.** For pairs of weights $w, \sigma$ that absolutely continuous with respect to Lebesgue measure, the norm inequality $\|P_+ (\sigma f)\|_w \leq \|f\|_\sigma$ holds if and only if the pair of weights satisfy the Poisson $A_2$ condition (1.4), and these testing inequalities hold, uniformly over all intervals $I$, for a finite positive constant $P$,

$$\int_I |P_+ (\sigma 1_I)|^2 w(dx) \leq P^2 \sigma(I), \quad \int_I |P_+ (w 1_I)|^2 \sigma(dx) \leq P^2 w(I).$$

One must be sure that the $A_2$ inequality is necessary from the norm inequality. As it suffices to test real-valued functions, the real-variable proof given here will suffice. This in particular shows that for the densities of the weights, $\sigma(x) \cdot w(x) \leq A_2$, for a.e.$x$. Thus, the identity part of the norm, and testing, inequalities are trivial. The remaining parts just concern the Hilbert transform, so one can use the main result.

If one is interested in the Sarason question for functions $f, g$ that are not outer, there is no simple reduction to the two weight inequality for the Hilbert transform, and the problem is quite subtle, as the role of the multiplier $P_+ M_\pi$ is more involved than that of just a weight.

**12.2. Model Spaces.** For a probability measure $\sigma$ on $T$, define a holomorphic function $\theta$ on $D$ by the Poisson integral

$$\frac{1}{1 - \theta(z)} := \int_T \frac{1}{1 - z\zeta} \sigma(d\zeta).$$

This is an inner function: A holomorphic map of $D$ to itself which is unimodular a.e. on $T$. Also, $\theta(0) = 1$. (The measure $\sigma$ is a Clark measure for $\theta$, frequently written as $\sigma_1$.)

The shift operator $Sf(z) = zf(z)$ on $H^2$ has invariant subspace $\theta H^2 = \{0 \theta : f \in H^2\}$, whence $K_\theta := H^2 \ominus \theta H^2$ is invariant for $S^*$. Beurling’s theorem states that every invariant subspace for $S^*$ is of this form. The model operator is $S_\theta := P_0 S$, where $P_0$ is the orthogonal projection from $H^2$ onto
Remarkably, subject to mild conditions, every contractive operator on a Hilbert space is unitarily equivalent to a properly chosen $S_{\theta}$. For this, and other reasons, properties of the $K_{\theta}$ spaces have broad significance.

The spaces $K_{\theta}$ and $L^2(\sigma)$ are unitarily equivalent, with the unitary map from $f \in L^2(\sigma)$ to $F \in K_{\theta}$ given by

$$F(z) = (1 - \theta(z)) \int_{\mathbb{T}} \frac{f(\zeta)}{1 - z\zeta} \sigma(d\zeta).$$

One is interested in those measures $\mu$ on $\mathbb{T}$ for which the natural embedding operator is bounded from $K_{\theta}$ to $L^2(\mu)$, namely, is it the case that $\|F\|_\mu \lesssim \|F\|_{K_{\theta}}$. We see that this bound is equivalent to

$$\int_{\mathbb{T}} \int_{\mathbb{T}} \left| \frac{f(\zeta)}{1 - z\zeta} \sigma(d\zeta) \right|^2 |1 - \theta(z)|^2 \mu(dz) \lesssim \|f\|^2_2.$$

That is, the question is equivalent to a two weight inequality for the Hilbert transform on $\mathbb{T}$.

From this perspective, one can lift counterexamples concerning the two weight Hilbert transform to those for embedding operators, which is the tactic of [37], from which we have taken this condensed presentation. A characterization of the embedding question can be read off from our main theorem. (But note that Clark measure is on $\mathbb{T}$, by definition, and the second measure $\mu$ is constrained to be supported on $\mathbb{T}$, whereas the disk would be the natural assumption.)

This subject is profound. The model spaces are also important to spectral theory, and the subject of rank one perturbations of a unitary operator. In spectral theory, it is important to understand the structure of the unitary operator that sends the Hilbert space to into $L^2$ of the spectral measure. Weighted Hilbert transforms arise therein. See for instance [39], which uses the example of Nazarov showing that the $A_2$ condition is not sufficient for the boundedness of the Hilbert transform. Also see [27].

We point the interested readers to [38, 43], and the many citations therein for more information about these subjects.

12.3. de Branges Spaces. We recall the setting of [3, 4]. For a sequence of distinct points $\Gamma = \{\gamma_n\} \subset \mathbb{C}$ and a sequence of positive numbers $\nu = \{\nu_n\}$ consider the Cauchy transform

$$H_{(\Gamma, \nu)} : \alpha = \{a_n\} \mapsto \sum_{n : z \neq \gamma_n} \frac{a_n \nu_n}{z - \gamma_n}.$$ 

This is well defined for $\alpha \in \ell^2_\nu$ and $z \in \Omega$, defined by

$$\Omega := \left\{ z \in \mathbb{C} : \sum_{n : z \neq \gamma_n} \frac{\nu_n}{|z - \gamma_n|^2} < \infty \right\}.$$

Call $\mathcal{H}(\Gamma, \nu)$ the space of functions analytic on $\Omega$ given by the image of $\ell^2_\nu$ under $H_{(\Gamma, \nu)}$. For appropriate choices of $(\Gamma, \nu)$, these Hilbert spaces have deep connections to analytic function spaces. For instance, the reproducing kernels of $\mathcal{H}(\Gamma, \nu)$ are

$$k_z(\zeta) := \sum_n \frac{\nu_n}{(z - \gamma_n)(\zeta - \gamma_n)}, \quad z \in \Omega.$$

And, many natural questions, such as the structure of frames of reproducing kernels for $\mathcal{H}(\Gamma, \nu)$, require knowledge about the two weight inequality for the Cauchy transform. For instance, the main real-variable result in [3] is a characterization of a two weight inequality, but under the requirement that both measures...
be a sum of point masses on sparse collections of points. This yields interesting results in the setting of de Branges spaces.

The definition of $\mathcal{H}(\Gamma,\nu)$ provides just one possible representation of a de Branges space, a class of Hilbert spaces with remarkable properties. The standard reference for them is [9]. Beginning from the works of Sarason [47], they have become an essential part of subject of analytic function spaces.

REFERENCES

[1] A. B. Aleksandrov, Isometric embeddings of co-invariant subspaces of the shift operator, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 232 (1996), no. Issled. po Linein. Oper. i Teor. Funktsii. 24, 5–15, 213 (Russian, with English and Russian summaries); English transl., J. Math. Sci. (New York) 92 (1998), no. 1, 3543–3549.

[2] Alexandru Aleman, Sandra Pott, and Maria Carmen Reguera, Sarason Conjecture on the Bergman space, available at http://www.arxiv.org/abs/1304.1750.

[3] Yurii Belov, Tesfa Y. Mengestie, and Kristian Seip, Discrete Hilbert transforms on sparse sequences, Proc. Lond. Math. Soc. (3) 103 (2011), no. 1, 73–105.

[4] ______, Unitary discrete Hilbert transforms, J. Anal. Math. 112 (2010), 383–393.

[5] Lennart Carleson, On convergence and growth of partial sums of Fourier series, Acta Math. 116 (1966), 135–157.

[6] Carme Cascanet, Joaquin M. Ortega, and Igor E. Verbitsky, On $L^p$-$L^q$ trace inequalities, J. London Math. Soc. (2) 74 (2006), no. 2, 497–511.

[7] David Cruz-Uribe, The invertibility of the product of unbounded Toeplitz operators, Integral Equations Operator Theory 20 (1994), no. 2, 231–237.

[8] David Cruz-Uribe, Jose Martell, and Carlos Pérez, A note on the off-diagonal Muckenhoupt-Wheeden conjecture (2012), available at http://www.arxiv.org/abs/1203.5906.

[9] Louis de Branges, Hilbert spaces of entire functions, Prentice-Hall Inc., Englewood Cliffs, N.J., 1968.

[10] Charles Fefferman, Pointwise convergence of Fourier series, Ann. of Math. (2) 98 (1973), 551–571.

[11] Richard Hunt, Benjamin Muckenhoupt, and Richard Wheeden, Weighted norm inequalities for the conjugate function and Hilbert transform, Trans. Amer. Math. Soc. 176 (1973), 227–251.

[12] Tuomas P. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, Ann. of Math. (2) 175 (2012), no. 3, 1473–1506.

[13] Tuomas Hytönen, On Petermichl’s dyadic shift and the Hilbert transform, C. R. Math. Acad. Sci. Paris 346 (2008), no. 21-22, 1133–1136 (English, with English and French summaries).

[14] Tuomas P. Hytönen, The $A_2$ theorem: Remarks and complements (2012), available at http://www.arxiv.org/abs/1212.3840.

[15] Tuomas P. Hytönen, Michael T. Lacey, Henri Martikainen, Tuomas Orponen, Maria Carmen Reguera, Eric T. Sawyer, and Ignacio 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.

[16] Michael T. Lacey, Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization, Part II, Submitted (2013), available at http://www.arxiv.org/abs/1301.4663.

[17] Michael T. Lacey, Stefanie Petermichl, and Maria Carmen Reguera, Sharp $A_2$ inequality for Haar shift operators, Math. Ann. 348 (2010), no. 1, 127–141.

[18] Michael T. Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, Two Weight Inequalities for Discrete Positive Operators, Submitted (2009), available at http://www.arxiv.org/abs/0911.3437.

[19] ______, A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure, Anal. PDE 5 (2012), no. 1, 1–60.

[20] ______, A Two Weight Inequality for the Hilbert transform Assuming an Energy Hypothesis, J Funct Anal 263 (2012), 305–363.

[21] Michael T. Lacey, Eric T. Sawyer, Chun-Yun Shen, and Ignacio Uriarte-Tuero, The two weight inequality for the Hilbert transform, coronas, and energy conditions (2011), available at http://www.arxiv.org/abs/1108.2319.

[22] ______, Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization, Part I, Submitted, available at http://www.arxiv.org/abs/1201.4319v9.

[23] ______, Two Weight Inequality for the Hilbert Transform: Parallel Corona, Submitted.
The Bellman function, the two-weight Hilbert transform, and embeddings of the model

F. Nazarov and A. Volberg,

Linear resolvent growth of rank one perturbation of a unitary operator does not imply its

Nikolai Nikolski and S. Treil,

N. K. Nikol’skiĭ,

Donald Sarason,

Exposed points in $L^p$,

Stefanie Petermichl,

Maria Carmen Reguera and Christoph Thiele,

The Hilbert transform does not map

Nikolai Nikolski, eds.), Lecture Notes in Mathematics, vol. 1573, Springer-Verlag, Berlin, 1994, pp. 318–319.

TWO WEIGHT INEQUALITIES FOR INDIVIDUAL HAAR MULTIPLIERS AND OTHER WELL LOCALIZED OPERATORS

The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical

S. Petermichl,

On Muckenhoupt-Wheeden conjecture

Aleksandrov-Clark measures

Two weight function norm inequalities for the Hardy-Littlewood maximal function and the Hilbert transform, Studia Math. 55 (1976), no. 3, 279–294.

Benjamin Muckenhoupt and Richard L. Wheeden, Two weight function norm inequalities for the Hardy-Littlewood maximal function and the Hilbert transform, Studia Math. 55 (1976), no. 3, 279–294.

F. Nazarov, A counterexample to Sarason’s conjecture, Preprint, MSU (1997), available at

Two weight inequalities for individual Haar multipliers and other well localized operators, Math. Res. Lett. 15 (2008), no. 3, 583–597.

F. Nazarov and A. Volberg, The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces $K_0$, J. Anal. Math. 87 (2002), 385–414. Dedicated to the memory of Thomas H. Wolff.

N. K. Nikol’skiĭ, Treatise on the shift operator, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 273, Springer-Verlag, Berlin, 1986. Spectral function theory; With an appendix by S. V. Hruščev [S. V. Khrushčëv] and V. V. Peller; Translated from the Russian by Jaak Peetre.

Nikolai 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. Dedicated to the memory of Thomas H. Wolff.

Stefanie Petermichl, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol, C. R. Acad. Sci. Paris Sér. I Math. 330 (2000), no. 6, 455–460 (English, with English and French summaries).

S. Petermichl, The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical $A_p$ characteristic, Amer. J. Math. 129 (2007), no. 5, 1355–1375.

Carlos Pérez, S. Treil, and A. Volberg, On $A_2$ conjecture and Corona decomposition of weights (2010), available at http://arxiv.org/abs/1006.2630.

Alexei Poltoratski and Donald Sarason, Aleksandrov-Clark measures, Recent advances in operator-related function theory, Contemp. Math., vol. 393, Amer. Math. Soc., Providence, RI, 2006, pp. 1–14.

Maria Carmen Reguera, On Muckenhoupt-Wheeden conjecture, Adv. Math. 227 (2011), no. 4, 1436–1450.

Maria Carmen Reguera and James Scurry, On joint estimates for maximal functions and singular integrals on weighted spaces, Proc. Amer. Math. Soc. 141 (2013), no. 5, 1705–1717.

Maria Carmen Reguera and Christoph Thiele, The Hilbert transform does not map $L^1(Mw)$ to $L^{1,\infty}(w)$, Math. Res. Lett. 19 (2012), no. 1, 1–7.

Donald Sarason, Exposed points in $H^1$, I, The Goebberg anniversary collection, Vol. II (Calgary, AB, 1988), Oper. Theory Adv. Appl., vol. 41, Birkhäuser, Basel, 1989, pp. 485–496.

Donald Sarason, Products of Toeplitz operators, Linear and complex analysis. Problem book 3. Part I (V. P. Havin and N. K. Nikolski, eds.), Lecture Notes in Mathematics, vol. 1573, Springer-Verlag, Berlin, 1994, pp. 318–319.

Eric T. Sawyer, A characterization of a two-weight norm inequality for maximal operators, Studia Math. 75 (1982), no. 1, 1–11.

Eric T. Sawyer, Ignacio Uriarte-Tuero, and Chun-Yen Shen, A two weight theorem for fractional singular integrals: an expanded version, available at http://www.arxiv.org/abs/1302.5093.
[53] E. Sawyer and R. L. Wheeden, Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces, Amer. J. Math. 114 (1992), no. 4, 813–874.

[54] J. Scurry, A Characterization of Two-Weight Inequalities for a Vector-Valued Operator (2010), available at http://www.arxiv.org/abs/1007.3089.

[55] H. Tanaka, A characterization of two weight trace inequalities for positive dyadic operators in the upper triangle case (2013), available at http://www.arxiv.org/abs/1302.4164.

[56] Sergei Treil, A remark on two weight estimates for positive dyadic operators (2012), available at http://www.arxiv.org/abs/1201.1455.

[57] A. L. Vol’berg, Thin and thick families of rational fractions, Complex analysis and spectral theory (Leningrad, 1979/1980), Lecture Notes in Math., vol. 864, Springer, Berlin, 1981, pp. 440–480.

[58] A. Volberg, Calderón-Zygmund capacities and operators on nonhomogeneous spaces, CBMS Regional Conference Series in Mathematics, vol. 100, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2003.

School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA
E-mail address: lacey@math.gatech.edu