WEIGHTED ALPERT WAVELETS

Structure and Stability of Weighted Alpert Wavelets

By Fletcher Gates, B.Math, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2025) Hamilton, Ontario (Mathematics)

TITLE: Structure and Stability of Weighted Alpert Wavelets

AUTHOR: Fletcher Gates, B.Math, M.Sc.

SUPERVISORS: Dr. Eric Sawyer, Dr. Scott Rodney

NUMBER OF PAGES: **vii, 86**

Abstract

In this thesis we present a number of results concerning Alpert wavelet bases for $L^2(\mu)$, with μ a locally finite positive Borel measure on $\mathbb{R}^n$. Alpert wavelets generalize Haar wavelets while retaining their orthonormality, telescoping, and moment vanishing properties. We show that the properties of such a basis are determined by the geometric structure of μ; in particular they are the result of linear dependences in $L^2(\mu)$ among the functions from which the wavelets are constructed; this completes an investigation begun by Rahm, Sawyer, and Wick [14]. These dependences can be efficiently detected using a Gröbner basis algorithm, which provides enough information to determine the structure of any Alpert basis constructed on μ. We present a generalization of the usual Alpert wavelet construction, where the degree of moment vanishing is allowed to vary over the dyadic grid. We also show that Alpert bases in a doubling measure on $\mathbb{R}$ are stable under small translations of the underlying dyadic intervals, building on work by Wilson [19]. We conclude with a partial result toward the converse, showing that a class of non-doubling measures cannot have this stability property.

Acknowledgements

First and foremost I would like to thank my advisors Dr. Eric Sawyer and Dr. Scott Rodney. You have been unfailingly supportive of me, and have ensured that grad school is a memory I will treasure for the rest of my life. I am grateful for all the help you've given me, and I hope we are able to work together again in the future.

I would also like to thank my parents Irene and Meinrad. I am sure you did not originally expect to spend six years learning about what an Alpert wavelet is, but I am so glad that you did. The Covid years in particular would have been very lonely without your company. Thanks as well to my brother Conrad; I love you dearly, but I'm afraid that won't stop me from insisting that you and you alone address me as Dr. Gates.

I am grateful for the many students I have been fortunate enough to teach over the past six years. You are too many to name, but your kindness and your enthusiasm for learning has always bolstered my own. Without the opportunity to share my love of math with all you it is likely this thesis would never have been finished. My best wishes to all of you in your future endeavours.

My endless appreciation to my friends: Flora, Jenn, Kathleen, Brian, Steven, Ashley, Justin, and Lauren. The margins of this page are too small to tell a funny-yet-endearing anecdote about each of you, so instead let me summarize and say that you are all the best.

To Sarah, simply: I love you. The world is a more beautiful place for having you in it.

Contents

List of Figures **vii**

1 Introduction **1**

1.1 Wavelet Bases 1

1.2 Weighted Wavelet Bases 3

1.3 Stability of Haar Wavelets 4

1.4 Summary of Main Results 5

1.4.1 Chapter 3 5

1.4.2 Chapter 4 7

2 Background **8**

2.1 Measure Theory 8

2.2 Dyadic Grids 15

2.3 Haar Wavelets 17

2.4 Alpert Wavelets 22

2.5 Weighted Alpert Wavelets 24

2.6 Algebraic Geometry 28

3 Structure of Weighted Alpert Wavelets **31**

3.1 Dimensions of Alpert Spaces 31

3.2 Variable Alpert Bases 35

3.3 Structure Theorem for Polynomial Alpert Bases 38

3.3.1 Suuplemental Results 42

3.3.2 Demonstrative Examples 46

4 Stability of Weighted Alpert Wavelets **52**

4.1 Doubling Measures 52

4.2 Stability and Almost-Orthogonality 55

4.3 Stability of Alpert Wavelets 60

4.3.1 Stability in Higher Dimensions 70

4.4 Instability in Non-Doubling Measures 72

5 Conclusion **76**

5.1 Further Questions 76

A Completeness of Haar Wavelets **79**

B Buchberger's Algorithm **83**

Bibliography **85**

List of Figures

2.3.1 Example 2 partial sum with $k = 1$. 21

2.3.2 Example 2 partial sum with $k = 2$. 21

2.3.3 Example 2 partial sum with $k = 3$. 21

3.2.1 Relations among classes of wavelet basis 39

3.3.1 A point mass measure yielding accidental orthogonality in $\mathbb{R}^2$ 45

3.3.2 The twisted cubic in $\mathbb{R}^3$ 49

Chapter 1

Introduction

We begin with a preface: both our objects of study and our results are somewhat cumbersome to define. In the interest of giving the reader a gentle introduction, this chapter will use "high-level" but imprecise statements for all but a few simple terms. Chapter 2 defines the necessary concepts and notation to make each of these statements precise.

1.1 Wavelet Bases

The standard dyadic grid D^* on $\mathbb{R}$ is the set of all half-open intervals which, for each $k \in \mathbb{Z}$, have length 2^k and endpoints which are consecutive multiples of 2^k. The real line is thereby decomposed into intervals of length 2^k at each scale $k \in \mathbb{Z}$. The classical Haar basis for $L^2(\mathbb{R})$ consists of a mother wavelet

$$h^{[0,1)}(x) = \mathbf{1}_{[0,\frac{1}{2})} - \mathbf{1}_{[\frac{1}{2},1)}$$

and then for each dyadic interval $I \in D$ the corresponding Haar wavelet h^I is a translate and dilate of the mother wavelet so that h^I is supported on I and has L^2-norm 1.

The set of all Haar wavelets forms an orthonormal basis for $L^2(\mathbb{R})$ with some highly desirable properties:

- Each Haar wavelet is piecewise constant.
- Each Haar wavelet is supported on a dyadic interval.

- Projection onto the Haar basis within some finite resolution simplifies via a telescoping property.
- Each Haar wavelet is orthogonal to constants.

This final property is often referred to as a *moment vanishing condition.*

Haar wavelets and their study date back to 1910 with the work of Alfred Haar [10]. In the years since, wavelets have proven to be highly useful for signal processing and its associated applications. Of particular note is an advantage which wavelets enjoy: because each wavelet is compactly supported, a wavelet transform can capture localized phenomena more efficiently than a traditional Fourier transform.

For a discusson of Haar wavelets in the context of signal processing, and in part also because we expect most readers are not fluent in Haar's native German, we suggest the survey by Stankovic and Falkowski [17] as an introductory article. For readers desiring a thorough treatise of Haar wavelets and how they relate to Fourier analysis, we recommend the book by Pereyra and Ward [13].

In practice, Haar wavelets are often not the preferred choice of wavelet for modern applications. More sophisticated constructions are able to achieve the same desireable properties of Haar wavelets while also achieving some other goal. For example, Daubechies wavelets [5] satisfy the same orthogonality conditions as Haar wavelets (more, in fact) while also being continuous.

A generalization of Haar wavelets was given by Alpert in 1993 [3]. Alpert's construction is a multiwavelet basis, where each wavelet is orthogonal to all polynomials with degree less than some fixed k. This is achieved by making two concessions:

1. Where each Haar wavelet is piecewise constant, each Alpert wavelet is instead piecewise polynomial of degree less than k.
2. Where the Haar basis has one wavelet per dyadic interval, an Alpert basis has k wavelets per dyadic interval.

Moreover, Alpert showed that such a basis is underdetermined and gave an algorithm for constructing a basis where some of the wavelets are additionally orthogonal to some higher-order polynomials. The two above concessions are substantial for practical applications, though Alpert showed that in some cases the computational advantage of the extra orthogonality outweighs the increase in complexity of the basis.

1.2 Weighted Wavelet Bases

Building on Alpert's construction, Rahm, Sawyer, and Wick [14] showed that Alpert bases also exist in $L^2(\mu)$, where μ is any locally finite positive Borel measure on $\mathbb{R}^n$. Of particular interest is the fact that such a basis retains almost all of the useful properties of Haar and Alpert bases in Lebesgue measure. The most notable loss is that, in a measure which is not translation invariant, the basis functions are no longer all translates and dilates of a set of mother wavelets.

Rahm, Sawyer, and Wick also observed that, in certain measures, the number of wavelets needed for each dyadic cube is sometimes less than would be needed in Lebesgue measure. This is discussed in [14, section 3]; in particular it is demonstrated in one dimension that an interval I with non-zero measure can nevertheless have one or more of its associated Alpert wavelets be identically zero. It was this observation which first motivated the research that would eventually become this thesis.

The resulting Alpert bases have been used to prove a number of results regarding Calderón-Zygmund operators—see for example [2] by Alexis, Sawyer, and Uriarte-Tuero, where Alpert bases have yielded partial progress toward extending the David-Journé $T1$ theorem [6] from Lebesgue measure to pairs of doubling measures. The additional moment vanishing afforded by Alpert wavelets is used throughout that result, justifying their use over the much simpler Haar wavelets. This use also demonstrates a drawback of Alpert wavelets; singular integrals do not in general commute with multiplication by polynomials, which is not a problem when working with the piecewise constant Haar wavelets.

Very recently, Alpert wavelets were used by Sawyer to prove a probabilistic analogue of the Fourier extension conjecture [16]. Here Sawyer first deals with a major shortcoming of Alpert wavelets, that being the discontinuity of each wavelet at the boundary of its support. Sawyer constructs a family of smooth Alpert 'wavelets', which coincide with standard Alpert wavelets except on a small halo around each discontinuity—where they are made smooth via a convolution—and which also retain Alpert's moment vanishing properties. The resulting functions no longer form a basis for $L^2(\mathbb{R}^n)$, but do form a frame (see [4] for more on frames). Both continuity and higher-order moment vanishing are necessary for the main proof.

1.3 Stability of Haar Wavelets

In a 2017 paper [19], Wilson showed that a particular class of $L^2(\mathbb{R}^n)$ functions were "stable", in the following sense: if $\{h_j\}_{j\in J}$ is a family of such functions, and f is any function in $L^2(\mathbb{R}^n)$, then the projection of f onto $\{h_j\}_{j\in J}$ is close to f itself in the L^2-norm. Clearly this holds when $\{h_j\}_{j\in J}$ is an orthonormal basis, but Wilson showed that this still holds when the h_j's in an orthonormal basis undergo a small amount of perturbation (translation, dilation, skew, etc).

Of particular interest to us is that one-dimensional Haar bases perturbed by small translations are among the class of functions to which Wilson's theorem applies. This is a desireable property for wavelet bases to have in practical applications; it ensures that small errors in computation won't dramatically effect the end result. Wilson extends this stability result to more general classes of functions in $\mathbb{R}^n$. In chapter 4 we extend in a different direction: we show that any one-dimensional weighted Alpert basis is also stable provided that the measure is doubling.

Wavelet perturbations also appear in the landmark result of Nazarov, Treil, and Volberg [12] as a key part of the proof that a set of testing criteria is sufficient to confirm boundedness of the Hilbert transform. Here a small translation is applied uniformly to the entire dyadic grid, then by taking an average over all such translations Nazarov et al. are able to control the "bad" behaviour of functions near the discontinuities in a Haar wavelet. These grid translations do not satisfy the definition of perturbation we consider in chapter 4, but are foundational to the $T1$ theorem results described in the previous section and so may be of interest to the reader anyway.

1.4 Summary of Main Results

1.4.1 Chapter 3

Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, let D be a dyadic grid on $\mathbb{R}^n$, and let U be a finite set of real-valued locally-integrable functions which contains the constant function 1. The associated Alpert basis for $L^2(\mu)$ consists of functions which are each supported on some cube $Q \in D$, are each in $\operatorname{Span} U$ when restricted to any child of Q, and are each orthogonal to all of U. This is a modest generalization of the bases described in the preceeding section, where U was specifically taken to be the set of all monomials less than some fixed degree. Given such an Alpert basis for $L^2(\mu)$, the number of Alpert wavelets needed for a given cube Q depends on both the set U and the underlying measure μ.

Theorem 3. *When constructing an Alpert basis, the number of wavelets needed for a given dyadic cube Q is the sum of the dimensions of $\operatorname{Span} U$ when restricted to each child of Q, minus the dimension of $\operatorname{Span} U$ when restricted to Q. Moreover, when applying extra orthogonality conditions to an Alpert basis, each additional condition is either achieved "for free" without affecting the basis, or otherwise is achieved by reducing the number of wavelets satisfying that condition by 1.*

The latter case of this conclusion explains the behaviour observed by Rahm, Sawyer, and Wick [14]. Their examples choose a measure μ and interval I where at least one child of I either had measure zero or a single point mass. The space of linear functions on that child will have either dimension zero or dimension one respectively, where in Lebesgue measure it would have dimension two, and so the Alpert wavelet basis for $L^2(\mu)$ will contain correspondingly fewer wavelets supported on I.

Prompted by the possibility of different intervals (or cubes in $\mathbb{R}^n$) having varying numbers of associated Alpert wavelets in a basis, we then considered the following question: is it possible to construct a basis which varies in this way without the variation being a direct consequence of the measure's geometry? The answer turns out to be yes, with a caveat:

Theorem 4. *For every dyadic cube Q let U_Q be a finite set of locally-integrable functions, and suppose that these sets obey the nesting condition $U_{P(Q)} \subseteq U_Q$, where $P(Q)$ denotes the parent of*

Q. Then an orthonormal basis for $L^2(\mu)$ can be constructed from the following three components:

1. *For every $Q \in D$, a set of Alpert wavelets on Q constructed from U_Q.*

2. *For every $Q \in D$, an orthonormal basis for the orthogonal complement of $\operatorname{Span} U_{P(Q)}$ inside $\operatorname{Span} U_Q$.*

3. *For every dyadic top $T \in \tau(D)$ (defined in section 2.2), an orthonormal basis for $\operatorname{Span} U_T$, where $U_T = \cap_{Q \subset T} U_Q$.*

This basis retains the same orthogonality properties as a standard Alpert basis.

A basis of this type allows for interpolation between the amount of orthogonality achieved and the number of wavelets needed for each cube. While this seems to capture the best elements of both Haar and Alpert bases, this improvement comes with the aforementioned caveat: the second component in the above list consists of the "leftover" terms when transitioning from cubes with more wavelets to cubes with fewer, and these leftover terms do not obey the Alpert wavelets' orthogonality conditions.

Section 3.3 provides a further refinement of theorem 3 for Alpert bases constructed from polynomials of bounded degree; this is the special case discussed by Rahm, Sawyer, and Wick in [14]. This result uses concepts from algebraic geometry, defined in section 2.6.

Theorem 5. *Let G be a Gröbner basis for the ideal of all polynomials on Q which vanish outside a set of μ-measure zero. The set of all monomials on Q which are not a multiple of any leading term in G is a maximal linearly independent set.*

By Theorem 3, knowing the dimensions of these polynomial spaces is sufficient to determine the number of Alpert wavelets needed for a given cube. This result has additional utility in that a single Gröbner basis calculation simultaneously provides the dimensions of these polynomial spaces for every choice of degree. This also provides a simple recipe for constructing examples where an Alpert space has less than the full dimension; μ when restricted to a cube Q should be supported inside the solution set to a set of polynomials with degrees no higher than the Alpert basis.

1.4.2 Chapter 4

Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$ and B be an orthonormal basis for $L^2(\mu)$. We consider perturbations of the basis elements in B—see definition 18—and we say that B is stable under a perturbation if

$$\sum_{b \in B} \langle f, b^* \rangle b^* \approx f$$

in the $L^2(\mu)$ norm, where b^* is the perturbation of b.

Theorem 8. *It suffices to test whether a series of inner products between the elements of B and the perturbed elements of B converges to a bound controlled by the magnitude of the perturbation. If it does, and if that bound vanishes as the magnitude of the perturbation tends to zero, then B is stable under that perturbation.*

This result adapts techniques developed by Wilson in [19] to our more general setting. While we only use this result in the context of Alpert wavelets, the full statement is evidently much more general. We have opted to present the more general version in case it happens to be of use in other areas.

Theorem 10. *If μ is a doubling measure on $\mathbb{R}$ and B is any polynomial Alpert basis for $L^2(\mu)$, then B satisfies the criterion from Theorem 8 and consequently is stable under small translations of the underlying dyadic intervals.*

We emphasize that μ is here taken to be specifically a one-dimensional doubling measure. It turns out that the condtion in Theorem 8 is sufficient to imply stability, but not necessary. In section 4.3.1 we show that the Haar basis in $\mathbb{R}^2$ does not satisfy this criterion, but is nevertheless known to be stable through work by Wilson [19, section 3].

Our final topic considers whether the above stability result holds in the opposite direction, or in other words whether any measure which satisfies our stability criterion must necessarily be doubling. This question turns out to be somewhat nuanced; a given non-doubling measure may be stable under a given perturbation for some dyadic grids and not for others. We conjecture that a measure which is stable under small translations for *all* dyadic grids must be doubling. Although we have not found a proof for this result in full, we do show that a substantial class of non-doubling measures cannot have this stability property for all dyadic grids.

Chapter 2

Background

If W is a finite-dimensional vector space with dimension n, we will refer to a subspace of dimension $n-1$ as a *hyperplane*. We define a *cube* $Q \in \mathbb{R}^n$ to be specifically oriented with sides parallel to the coordinate hyperplanes, i.e. a set of the form

$$Q = \{x \in \mathbb{R}^n : a_i \leq x_i < b_i, i = 1, \dots, n\} = \prod_{i=1}^{n} [a_i, b_i)$$

where each $a_i, b_i \in \mathbb{R}$ is constant and where there is a constant $l(Q) > 0$ such that

$$b_1 - a_1 = b_2 - a_2 = \cdots = b_n - a_n = l(Q)$$

We call $l(Q)$ the side length of Q. The cube mQ is the cube which has the same center as Q and has side length $m \cdot l(Q)$. We will refer to an interval I rather than a cube Q when working in dimension $n = 1$. In this case we will write I_l and I_r to denote the left and right halves of I respectively.

2.1 Measure Theory

This section contains a brief introduction to the concepts in measure theory which are needed to read this work; for a more detailed treatise see [18, Chapters 1 & 2]. Our definitions here will be given for an arbitrary set X, but in practice we will only ever take X to be some Euclidean space $\mathbb{R}^n$.

Definition 1 (σ-algebra)**.** *Let X be a set. A σ-algebra A on X is a collection of subsets of X which satisfies:*

1. *$\varnothing \in A$.*
2. *A is closed under complements: $S^c \in A$ for every $S \in A$.*
3. *A is closed under countable unions: $\bigcup_{i=1}^{\infty} S_i \in A$ whenever $S_i \in A$ for all $i \in \mathbb{N}$.*

As an immediate consequence of properties 2 and 3, σ-algebras are also closed under countable intersections. The pair (X, A) is called a *measurable space*.

Definition 2 (Measure)**.** *Let X be a set and let A be a σ-algebra on X. A measure μ is function $\mu : A \to \mathbb{R} \cup \{\infty\}$ which satisfies:*

1. *$\mu(\varnothing) = 0$.*
2. *For all $S \in A$, $\mu(S) \geq 0$.*
3. *For all countable collections $\{S_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in A,*

$$\mu\left(\bigcup_{i=1}^{\infty} S_i\right) = \sum_{i=1}^{\infty} \mu(S_i).$$

The triple (X, A, μ) is called a *measure space*, and the members of A are the *measurable sets* (or the μ-measurable sets, when we wish to be precise). Given a measurable set S, we will refer to $\mu(S)$ as the measure of S. There exists a notion of *signed measure*, which is as above except that sets are also allowed to have negative measure, but we do not consider signed measures in this work.

Definition 3 (Topology)**.** *Let X be a set. A topology τ on X is a collection of subsets of X, called open sets, which satisfies:*

1. *$\varnothing \in \tau$ and $X \in \tau$.*
2. *For any index set I, finite or infinite, $\bigcup_{i \in I} S_i \in A$ whenever $S_i \in \tau$ for all $i \in I$.*

3. *For* $k \in \mathbb{N}$, $\bigcap_{i=1}^{k} S_i \in A$ *whenever* $S_i \in \tau$ *for all* $i = 1, \ldots, k$.

The pair (X, τ) is called a *topological space*. A familiar example is the set of all open intervals on $\mathbb{R}$. A topology τ on X gives rise to a natural σ-algebra on X:

Lemma 1. *Let* (X, τ) *be a topological space. There is a unique* σ*-algebra* B *on* X *such that*

1. $\tau \subseteq B$.

2. *If* A *is any* σ*-algebra on* X *which contains* τ*, then* $B \subseteq A$.

Proof. Define the set

$$B = \bigcap \{A : A \text{ is a } \sigma\text{-algebra on } X \text{ with } \tau \subseteq A\}.$$

The power set $P(X)$ is a σ-algebra on X which contains τ, so B is non-empty. Suppose that $S \in B$, so $S \in A$ for every σ-algebra A on X with $\tau \subseteq A$. By σ-algebra properties, $S^c \in A$ for every such A and consequently $S^c \in B$. An identical argument shows that B is also closed under countable unions, so B is a σ-algebra on X. The properties

1. $\tau \subseteq B$.

2. If A is any σ-algebra on X which contains τ, then $B \subseteq A$.

follow immediately from the definition of B. □

The above σ-algebra B is called the *Borel* σ*-algebra*. The elements of B are the *Borel sets*, and any measure μ defined over the measurable space (X, B) is likewise called a *Borel measure*.

Definition 4 (Locally finite measure)**.** *Let* (X, τ) *be a topological space and* A *be a sigma-algebra on* X *that contains* τ*. A measure* μ *defined on* (X, A) *is locally finite if for every* $p \in X$ *there exists an open set* $S_p \in \tau$ *such that* $p \in S_p$ *and* $\mu(S_p) < \infty$.

All measures we consider from this point on are locally finite positive Borel measures on $\mathbb{R}^n$ for some $n \in \mathbb{N}$. Because we work exclusively in this context, we will not bother to specify the σ-algebras or topologies in our notation beyond this section.

Lemma 2. *Let τ be the standard topology on $\mathbb{R}^n$, and let μ be a locally finite positive Borel measure. For every cube $Q \subset \mathbb{R}^n$, Q is μ-measurable.*

Proof. Let $Q = \prod_{i=1}^n [a_i, b_i)$. Fix $\epsilon > 0$, and we can write

$$Q = \prod_{i=1}^n [a_i, b_i] \cap \prod_{i=1}^n (a_i - \epsilon, b_i).$$

The first product is the complement of an open set and the second product is an open set, so Q is a Borel set and therefore is μ-measurable. □

We will work heavily with function spaces defined over cubes; this is why we restrict ourselves to considering Borel measures.

Definition 5 (Measurable function)**.** *Let (X_1, A_1) and (X_2, A_2) be two measurable spaces. A function $f : X_1 \to X_2$ is measurable if for every $S \in A_2$ the preimage of S under f is in A_1, that is*

$$f^{-1}(S) = \{x \in X : f(x) \in S\} \in A_1.$$

We will only consider functions $f : \mathbb{R}^n \to \mathbb{R}$ for some $n \in \mathbb{N}$, so we will refer to μ-measurable functions f and trust that the underlying measure space is clear from context.

The primary motivation for the above definitions is to allow us to construct the Lebesgue integral. Recall that the Riemann integral can be loosely described as follows: given a function f, partition the domain of integration and approximate the area under f with vertical rectangles separated by the partition. The Lebesgue integral instead partitions the range of f, so the area under f is approximated by horizontal rectangles. Provided that f is measurable, each of these rectangles will have a "footprint" in the domain which is measurable.

To make the above idea precise, we define the Lebesgue integral in stages. First, let μ be a locally finite positive Borel measure on $\mathbb{R}^n$ and let $S \subseteq \mathbb{R}^n$ be a μ-measurable set. We will write $\mathbf{1}_S$ to mean the indicator function on S, that is

$$\mathbf{1}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}.$$

Note that any preimage of $\mathbf{1}_S$ is either all of S, all of S^c, all of $\mathbb{R}^n$, or empty. These are all measurable sets, so $\mathbf{1}_S$ is a measurable function. We define the Lebesgue integral of $\mathbf{1}_S$:

$$\int_{\mathbb{R}^n} \mathbf{1}_S(x)\, d\mu(x) = \mu(S).$$

This integral is allowed to take the value ∞.

Next, a *simple function* g is any finite linear combination of indicator functions

$$g = \sum_{i=1}^{m} a_i \mathbf{1}_{S_i}, \quad a_i \in \mathbb{R}$$

where the S_i are disjoint measurable sets. The range of a simple function is a finite set, so simple functions are measurable by the same argument as for indicator functions. For a simple function $g = \sum_{i=1}^{m} a_i \mathbf{1}_{S_i}$ with $a_i \geq 0$ for all $i = 1, \ldots, m$ we define the Lebesgue integral of g:

$$\int_{\mathbb{R}^n} g(x)\, d\mu(x) = \sum_{i=1}^{m} a_i \mu(S_i)$$

where again this sum might yield ∞. To verify that this is well-defined, suppose that

$$g = \sum_{i=1}^{m_1} a_i \mathbf{1}_{S_i} \quad \text{and} \quad g = \sum_{j=1}^{m_2} b_j \mathbf{1}_{T_j}$$

are two different representations of g. For any $x \in \mathbb{R}^n$, if $x \in S_i$ and $x \in T_j$ then $a_i = b_j$. Then fix $k \in \mathbb{R}$; the disjoint union of all S_i such that $a_i = k$ is equal to the disjoint union of all T_j such that $b_j = k$. Since measures are additive under disjoint unions, by letting k vary over the range of g we conclude

$$\sum_{i=1}^{m_1} a_i \mu(S_i) = \sum_{j=1}^{m_2} b_j \mu(T_j)$$

so the Lebesgue integral of a simple function is well-defined. Likewise if $T \subset \mathbb{R}^n$ is any measurable set, we define

$$\int_T g(x)\, d\mu(x) = \int_{\mathbb{R}^n} (\mathbf{1}_T g)(x)\, d\mu(x) = \sum_{i=1}^{m} a_i \mu(S_i \cap T).$$

Now let $f : \mathbb{R}^n \to [0, \infty]$ be a non-negative measurable function. We define the Lebesgue integral of f:

$$\int_{\mathbb{R}^n} f(x)\,d\mu(x) = \sup_g \left\{ \int_{\mathbb{R}^n} g(x) : 0 \leq g \leq f \text{ and } g \text{ is simple} \right\}.$$

As usual, this supremum may take value ∞.

Finally let $f : \mathbb{R}^n \to [-\infty, \infty]$ be any measurable function. Write $f = f^+ - f^-$ where

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) > 0 \\ 0 & \text{otherwise} \end{cases},$$

$$f^-(x) = \begin{cases} -f(x) & \text{if } f(x) < 0 \\ 0 & \text{otherwise} \end{cases},$$

so both f^+ and f^- are non-negative measurable functions. We say the Lebesgue integral of f exists if at least one of $\int_{\mathbb{R}^n} f^+(x)\,d\mu(x)$ and $\int_{\mathbb{R}^n} f^-(x)\,d\mu(x)$ is finite, in which case we define

$$\int_{\mathbb{R}^n} f(x)\,d\mu(x) = \int_{\mathbb{R}^n} f^+(x)\,d\mu(x) - \int_{\mathbb{R}^n} f^-(x)\,d\mu(x).$$

We say that f is *Lebesgue integrable* if $\int_{\mathbb{R}^n} |f(x)|\,d\mu(x) < \infty$. If $T \subset \mathbb{R}^n$ is a measurable set, we define

$$\int_T f(x)\,d\mu(x) = \int_{\mathbb{R}^n} (\mathbf{1}_T f)(x)\,d\mu(x).$$

Now that we have the Lebesgue integral defined in full, we turn to the function spaces it creates. Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$. The space $L^2(\mu)$ is the space of μ-measurable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^n} |f(x)|^2\,d\mu(x) < \infty.$$

Two functions $f, g \in L^2(\mu)$ are identified if the set $\{x \in \mathbb{R}^n : f(x) \neq g(x)\}$ is measurable and has measure zero, in which case we write $f = g$ μ-*almost everywhere*. For $f \in L^2(\mu)$ we have the norm

$$\|f\|_2 = \left(\int_{\mathbb{R}^n} |f(x)|^2\,d\mu(x) \right)^{\frac{1}{2}},$$

and for $f, g \in L^2(\mu)$ we have the inner product

$$\langle f, g \rangle = \int_{\mathbb{R}^n} f(x)g(x)\, d\mu(x).$$

For any cube Q in $\mathbb{R}^n$, denote by $L^2_Q(\mu)$ the space $\{\mathbf{1}_Q f : f \in L^2(\mu)\}$. Also let $L^2_{\text{loc}}(\mu)$ denote the space of locally square-integrable functions on $\mathbb{R}^n$ with respect to μ. Thus $L^2_Q(\mu) \subseteq L^2_{\text{loc}}(\mu)$ for any cube Q.

Lemma 3 (Schur Test)**.** *Let (X, μ) be a measure space with μ a positive measure. Let K be a non-negative measurable function on $X \times X$. Define the integral operator*

$$Tf(x) = \int_X K(x, y)f(y)\, d\mu(y), x \in X.$$

If there exists a constant $M > 0$ such that

$$\int_X K(x, y)\, d\mu(y) \leq M$$

for almost all $x \in X$, and

$$\int_X K(x, y)\, d\mu(x) \leq M$$

for almost all $y \in X$, then $\|T\| \leq M$ in $L^2(X, \mu)$.

Proof. Let $f \in L^2(X, \mu)$. By the Cauchy-Schwarz inequality we have

$$\begin{aligned} |Tf(x)|^2 &= \left| \int_X K(x, y)f(y)\, d\mu(y) \right|^2 \\ &\leq \left| \int_X K(x, y)\, d\mu(y) \right| \left| \int_X K(x, y)f(y)^2\, d\mu(y) \right| \\ &\leq M \left| \int_X K(x, y)f(y)^2\, d\mu(y) \right|. \end{aligned}$$

Integrating the above inequality with respect to x yields

$$\int_X |Tf(x)|^2\, d\mu(x) \leq M \int_X \left| \int_X K(x, y)f(y)^2\, d\mu(y) \right| d\mu(x).$$

The lefthand side of this inequality is exactly $\|Tf\|_2^2$, and in the righthand side we can exchange the order of integration using Fubini's Theorem [18, p. 279, Theorem 3.3]:

$$\begin{aligned}
\|Tf\|_2^2 &\leq M \int_X \left| \int_X K(x,y) f(y)^2 \, d\mu(y) \right| d\mu(x) \\
&= M \int_X \left| \int_X K(x,y) f(y)^2 \, d\mu(x) \right| d\mu(y) \\
&= M \int_X \left| \int_X K(x,y) \, d\mu(x) \right| f(y)^2 \, d\mu(y) \\
&\leq M \int_X M f(y)^2 \, d\mu(y) \\
&= M^2 \|f\|_2^2.
\end{aligned}$$

We conclude that $\|Tf\|_2 \leq M\|f\|_2$ for any $f \in L^2(X, \mu)$, as desired. □

There exist more sophisticated formulations of the Schur test, but this is sufficient for our purposes. In fact we will only need the special case where μ is the counting measure on a dyadic grid.

2.2 Dyadic Grids

Definition 6 (Dyadic Grid)**.** *A dyadic grid D on $\mathbb{R}^n$ is a set of half-open cubes with the following properties*

1. *Every cube $Q \in D$ has $l(Q) = 2^m$ for some $m \in \mathbb{Z}$.*
2. *Every cube $Q \in D$ with $l(Q) = 2^m$ is contained in some $R \in D$ with $l(R) = 2^{m+1}$.*
3. *For every $m \in \mathbb{Z}$, the subset of all cubes in D with length 2^m forms a partition of $\mathbb{R}^n$.*

The *standard dyadic grid* D^* on $\mathbb{R}^n$ is the unique dyadic grid where no cube has an interior that overlaps a coordinate hyperplane. Given a dyadic grid D and $Q \in D$ with side length 2^m, the set of *children* $C(Q)$ is the set of 2^n cubes in D contained within Q having length 2^{m-1}. Likewise the *parent* $P(Q)$ of a dyadic cube Q is the unique cube in D which contains Q and has length $2 \cdot l(Q)$.

Let D be a dyadic grid on $\mathbb{R}^n$ and $Q \in D$. We define the *tower* $\Gamma(Q)$ as the set of all cubes in D which contain Q. The *top* T of a tower $\Gamma(Q)$ is defined

$$T = \bigcup_{R \in \Gamma(Q)} R.$$

As an immediate consequence of the nesting property of towers, any two towers will either have the same top or two disjoint tops. This gives an equivalence relation on the towers in D, where two towers Γ_1 and Γ_2 are equivalent if their tops coincide. Since it is the tops in which we are primarily interested, we will simply refer to the unique set of tops arising from any choice of representatives from each equivalence class. Let $\tau(D)$ denote the set of unique tops in S.

Example 1. *The standard dyadic grid D^* on $\mathbb{R}^2$ has four unique tops, corresponding to the four quadrants in $\mathbb{R}^2$. In contrast, we can construct a dyadic grid D with only a single top as follows. Choose any square Q in $\mathbb{R}^2$; to be part of a dyadic grid, Q can have one of four possible parents corresponding to the four diagonal directions relative to the coordinate axes. If we construct the tower $\Gamma(Q)$ and at each level in the tower we cycle through those four diagonal directions in sequence, then the top of $\Gamma(Q)$ must extend infinitely in all directions.*

More generally, the standard dyadic grid D^* has exactly 2^n tops, corresponding to the orthants in $\mathbb{R}^n$, and these tops partition $\mathbb{R}^n$. While we have just seen that it is possible to have fewer than 2^n tops, the partition property holds in general.

Lemma 4. *For any dyadic grid D on $\mathbb{R}^n$, $\tau(D)$ forms a partition of $\mathbb{R}^n$ with no more than 2^n elements.*

Proof. Every dyadic cube has sides parallel to the coordinate hyperplanes, so every top is the product of n infinite intervals where each interval is either unbounded or has one finite upper or lower bound. Let T_1 and T_2 be two tops in $\tau(D)$ with

$$T_1 = \prod_{i=1}^{n} I_i \text{ and } T_2 = \prod_{i=1}^{n} J_i.$$

Suppose that for each $i \in 1, ..., n$, I_i and J_i extend infinitely it the same direction. Then T_1 and T_2 must have non-empty intersection, and consequently $T_1 = T_2$. By the contrapositive, if $T_1 \neq T_2$ then there must be at least one $i \in 1, ...n$ where I_i and J_i do not extend infinitely in the same direction, and there are at most 2^n ways to make such a choice. Lastly since every $x \in \mathbb{R}^n$ is contained in some dyadic cube, it is also contained in some top. Therefore $\tau(D)$ is a partition of $\mathbb{R}^n$ with no more than 2^n elements. $\square$

Informally, this proof observes that each of the 2^n oriented diagonals in $\mathbb{R}^n$ can belong to only one top. It is worth emphasizing that although tops are not themselves dyadic cubes, their boundaries still run parallel to the coordinate hyperplanes. Tops can therefore be thought of as "infinite dyadic cubes".

Remark 1. *The pattern in example 1 holds in general: any dyadic grid in $\mathbb{R}^n$ can be constructed by choosing a cube Q in $\mathbb{R}^n$ and constructing a tower $\Gamma(Q)$ above it. The choice of Q determines the grid for all scales smaller than $l(Q)$, and at each level in $\Gamma(Q)$ the choice among the 2^n possible parents of Q determines the grid at that scale.*

2.3 Haar Wavelets

Let D be a dyadic grid on $\mathbb{R}$. For each $I \in D$ the Haar function $h^I(x)$ is defined as:

$$h^I(x) = \frac{1}{\sqrt{l(I)}} \left(\mathbf{1}_{I_l}(x) - \mathbf{1}_{I_r}(x) \right).$$

This is the unique function (up to multiplication by -1) in $L^2(\mathbb{R})$ which has all of the following properties:

- h^I is supported on I.
- h^I is constant on each child of I.
- $\|h^I\|_2 = 1$.
- $\int_I h^I \, dx = 0$.

Each of these individually is a convenient property to have, but of greater significance is what this implies about the family of all Haar functions on D.

Theorem 1 (Haar Bases)**.** *The family of Haar functions $\{h^I\}_{I \in D}$ has the following properties:*

1. *$\{h^I\}_{I\in D}$ is an orthonormal basis for $L^2(\mathbb{R})$.*

2. *Projection on to $\{h^I\}_{I\in D}$ satisfies a telescoping property: for any $f \in L^2(\mathbb{R})$ and integers $m < n$,*

$$\sum_{I\in D:\, 2^{m+1}\le l(I)\le 2^n} \left\langle f, h^I\right\rangle h^I(x) = \sum_{I\in D:\, l(I)=2^m} \left\langle f, \frac{1}{2^m}\mathbf{1}_I\right\rangle \mathbf{1}_I(x) - \sum_{I\in D:\, l(I)=2^n} \left\langle f, \frac{1}{2^n}\mathbf{1}_I\right\rangle \mathbf{1}_I(x).$$

3. *Each h^I satisfies a moment vanishing condition:*

$$\int_I h^I(t)\,dt = 0.$$

Proof. Property 3 follows immediately from the definition of h^I. For property 2, let $I \in D$ and $f \in L^2(\mathbb{R})$. We have

$$\begin{aligned}
&\left\langle f, h^I\right\rangle h^I(x)\\
&= \frac{1}{l(I)}\int_I f(t)\cdot(\mathbf{1}_{I_l}(t)-\mathbf{1}_{I_r}(t))\,dt\cdot(\mathbf{1}_{I_l}(x)-\mathbf{1}_{I_r}(x))\\
&= \frac{1}{l(I)}\left(\int_{I_l} f(t)\,dt - \int_{I_r} f(t)\,dt\right)\cdot(\mathbf{1}_{I_l}(x)-\mathbf{1}_{I_r}(x))\\
&= \frac{1}{l(I)}\left(\int_{I_l} f(t)\,dt\cdot\mathbf{1}_{I_l}(x) + \int_{I_r} f(t)\,dt\cdot\mathbf{1}_{I_r}(x) - \int_{I_r} f(t)\,dt\cdot\mathbf{1}_{I_l}(x) - \int_{I_l} f(t)\,dt\cdot\mathbf{1}_{I_r}(x)\right).
\end{aligned}$$

We also have

$$\begin{aligned}
&\left\langle f, \frac{1}{l(I_l)}\mathbf{1}_{I_l}\right\rangle \mathbf{1}_{I_l}(x) + \left\langle f, \frac{1}{l(I_r)}\mathbf{1}_{I_r}\right\rangle \mathbf{1}_{I_r}(x) - \left\langle f, \frac{1}{l(I)}\mathbf{1}_I\right\rangle \mathbf{1}_I(x)\\
&= \frac{2}{l(I)}\int_{I_l} f(t)\,dt\cdot\mathbf{1}_{I_l}(x) + \frac{2}{l(I)}\int_{I_r} f(t)\,dt\cdot\mathbf{1}_{I_r}(x) - \frac{1}{l(I)}\int_I f(t)\,dt\cdot\mathbf{1}_I(x)\\
&= \frac{1}{l(I)}\left(2\int_{I_l} f(t)\,dt\cdot\mathbf{1}_{I_l}(x) + 2\int_{I_r} f(t)\,dt\cdot\mathbf{1}_{I_r}(x) - \int_I f(t)\,dt\cdot\mathbf{1}_I(x)\right).
\end{aligned}$$

Then applying the decomposition

$$\int_I f(t)\,dt\cdot\mathbf{1}_I(x) = \int_{I_l} f(t)\,dt\cdot\mathbf{1}_{I_l}(x) + \int_{I_r} f(t)\,dt\cdot\mathbf{1}_{I_r}(x) + \int_{I_r} f(t)\,dt\cdot\mathbf{1}_{I_l}(x) + \int_{I_l} f(t)\,dt\cdot\mathbf{1}_{I_r}(x),$$

we conclude

$$\left\langle f, h^I\right\rangle h^I(x) = \left\langle f, \frac{1}{l(I_l)}\mathbf{1}_{I_l}\right\rangle \mathbf{1}_{I_l}(x) + \left\langle f, \frac{1}{l(I_r)}\mathbf{1}_{I_r}\right\rangle \mathbf{1}_{I_r}(x) - \left\langle f, \frac{1}{l(I)}\mathbf{1}_I\right\rangle \mathbf{1}_I(x).$$

Next fix $m \in \mathbb{Z}$ and let $n = m + 1$. Adding the above equality over every interval of length 2^n gives

$$\sum_{I \in D:\, l(I)=2^n} \langle f, h^I \rangle h^I(x) = \sum_{I \in D:\, l(I)=2^{n-1}} \left\langle f, \frac{1}{2^{n-1}} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) - \sum_{I \in D:\, l(I)=2^n} \left\langle f, \frac{1}{2^n} \mathbf{1}_I \right\rangle \mathbf{1}_I(x).$$

Lastly we observe that when $n > m + 1$ the righthand sums telescope:

$$\begin{aligned}
&\sum_{i=m+1}^{n} \left(\sum_{I \in D:\, l(I)=2^{i-1}} \left\langle f, \frac{1}{2^{i-1}} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) - \sum_{I \in D:\, l(I)=2^i} \left\langle f, \frac{1}{2^i} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \right) \\
&= \sum_{i=m+1}^{n} \left(\sum_{I \in D:\, l(I)=2^{i-1}} \left\langle f, \frac{1}{2^{i-1}} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \right) - \sum_{i=m+1}^{n} \left(\sum_{I \in D:\, l(I)=2^i} \left\langle f, \frac{1}{2^i} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \right) \\
&= \sum_{I \in D:\, l(I)=2^m} \left\langle f, \frac{1}{2^m} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) + \sum_{i=m+2}^{n} \left(\sum_{I \in D:\, l(I)=2^{i-1}} \left\langle f, \frac{1}{2^{i-1}} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \right) \\
&\qquad - \sum_{i=m+1}^{n-1} \left(\sum_{I \in D:\, l(I)=2^i} \left\langle f, \frac{1}{2^i} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \right) - \sum_{I \in D:\, l(I)=2^n} \left\langle f, \frac{1}{2^n} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \\
&= \sum_{I \in D:\, l(I)=2^m} \left\langle f, \frac{1}{2^m} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) - \sum_{I \in D:\, l(I)=2^n} \left\langle f, \frac{1}{2^n} \mathbf{1}_I \right\rangle \mathbf{1}_I(x).
\end{aligned}$$

With this we arrive at the desired conclusion:

$$\sum_{I \in D:\, 2^{m+1} \le l(I) \le 2^n} \langle f, h^I \rangle h^I(x) = \sum_{I \in D:\, l(I)=2^m} \left\langle f, \frac{1}{2^m} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) - \sum_{I \in D:\, l(I)=2^n} \left\langle f, \frac{1}{2^n} \mathbf{1}_I \right\rangle \mathbf{1}_I(x).$$

For property 1, let $I, J \in D$. If $I = J$ then we have

$$\langle h^I, h^J \rangle = \int_I h^I(x)^2 \, dx = \frac{1}{l(I)} \int_I \mathbf{1}_I(x) \, dx = 1.$$

Otherwise if $I \neq J$ then either h^I and h^J have disjoint support, or one of h^I and h^J is supported inside a child of the other. In the former case we trivially have $\langle h^I, h^J \rangle = 0$, and in the latter case we have $\langle h^I, h^J \rangle = 0$ as a consequence of property 3. The set of Haar functions are therefore orthonormal.

To complete the proof it remains to show that the Haar functions are dense in $L^2(\mathbb{R})$. This is somewhat involved and uses techniques which are not needed to read this work, so in the interest of brevity we will omit it here. For those interested the full proof is given in appendix A. $\square$

The Haar basis has an additional useful property: all Haar functions are translates and dilates of one another. It is common to present Haar wavelets by first defining a *mother wavelet*, usually $h^{[0,1)}$, and then defining the other Haar functions as translates and dilates of the mother wavelet. However, unlike the three properties in Theorem 1, this relies on the translation-invariance of Lebesgue measure, a property which is not true in general for an arbitrary locally finite positive Borel measure.

For readers who have not worked with Haar wavelets before, we wish to address a common question: how can a set of integral-zero functions be a basis for $L^2(\mathbb{R})$, which clearly contains functions which do not have integral zero? Indeed, let $f \in L^2(\mathbb{R})$ and consider the following erroneous computation:

$$\int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} \sum_{I \in D} \langle f, h^I \rangle h^I(x)\,dx = \sum_{I \in D} \langle f, h^I \rangle \int_{\mathbb{R}} h^I(x)\,dx = 0.$$

The author confesses to making precisely this error as a student. The resolution to the mystery is that the second equality here is invalid; the integral does not commute with an infinite sum as it would a finite sum.

Example 2. *To see this resolution in action, let* $f = \mathbf{1}_{[0,1)}$. *We have* $\langle f, h^I \rangle \neq 0$ *precisely when* $I = [0, 2^k)$ *for* $k = 1, 2, 3...$, *and so*

$$\sum_{I \in D} \langle f, h^I \rangle h^I = \sum_{k=1}^{\infty} 2^{-k/2} h^{[0,2^k)}.$$

Since the smallest interval in this sum is $[0, 2)$, *every Haar function contributes some positive amount at values* $x \in [0, 1)$; *this allows the sum to converge to the desired indicator function. By contrast, each Haar function contributes some unwanted negative amount at* x*-values greater than 1 and these are canceled out by Haar functions later in the sum. In this way the error between each finite sum and* f *is "pushed out" toward infinity. The result is*

$$\lim_{m \to \infty} \sum_{k=1}^{m} 2^{-k/2} h^{[0,2^k)}(x) = 1 \text{ for } x \in [0,1), \text{ and}$$
$$\lim_{m \to \infty} \sum_{k=1}^{m} 2^{-k/2} h^{[0,2^k)}(x) = 0 \text{ for } x \notin [0,1).$$

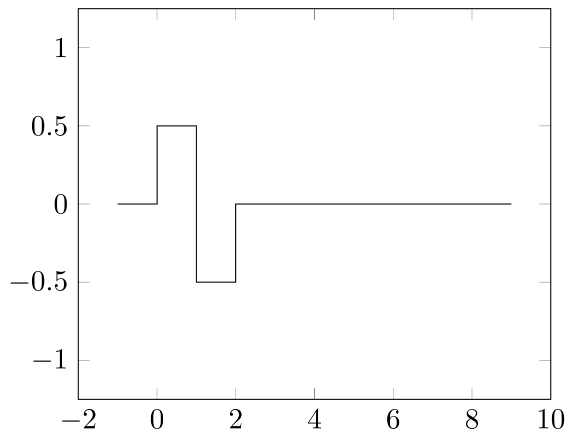

Figure 2.3.1: Example 2 partial sum with $k = 1$.

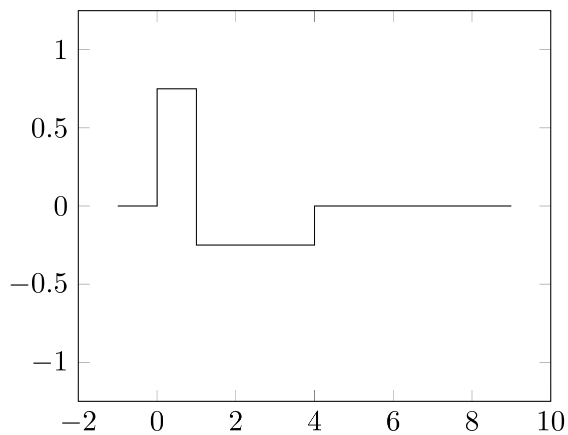

Figure 2.3.2: Example 2 partial sum with $k = 2$.

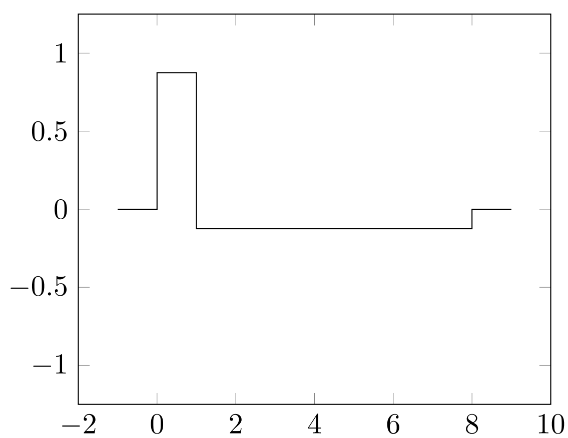

Figure 2.3.3: Example 2 partial sum with $k = 3$.

2.4 Alpert Wavelets

We say a function $f(x)$ has k-degree moment vanishing if

$$\int_{\mathbb{R}} f(x) \cdot x^k dx = 0.$$

This generalizes the moment vanishing condition satisfied by Haar functions, which have moment vanishing only for $k = 0$. In Alpert's construction [3], moment vanishing conditions up to degree $k-1$ are achieved by replacing each Haar function h^I with a family of k Alpert functions $\{a_j^I\}_{j=1,\ldots,k}$. Where each Haar function is piecewise constant on an interval I, each Alpert wavelet is piecewise polynomial of degree less than k. Bases where each dyadic cube has multiple associated wavelets are called *multiwavelet bases*.

The resulting system of equations is underdetermined. There are $\binom{k}{2}$ extra degrees of freedom which can be used to impose higher-degree moment vanishing conditions. Alpert's construction gives a basis with a "triangular tower" of additional moment vanishing properties:

- All k functions have $(k-1)$-degree moment vanishing.
- $k-1$ functions have k-degree moment vanishing.
- $k-2$ functions have $(k+1)$-degree moment vanishing.

  $\vdots$
- 2 functions have $(2k-3)$-degree moment vanishing.
- 1 function has $(2k-2)$-degree moment vanishing.

To see this let $I \in D$ and $k \in \mathbb{N}$. The family of Alpert functions $a_j^I(x)$, $j = 1, \ldots, k$ is constructed as follows:

1. For $j = 1, \ldots, k$ define the functions f_j^1:

$$f_j^1(x) = x^{j-1} \cdot \mathbf{1}_{I_l} - x^{j-1} \cdot \mathbf{1}_{I_r}.$$

2. Apply the Gram Schmidt algorithm to the set $\{1, x, \ldots, x^{k-1}, f_1^1, \ldots, f_k^1\}$. The $k-1$ outputs associated to $\{f_1^1, \ldots, f_k^1\}$ are each orthogonal to $1, x, \ldots, x^{k-1}$; label these outputs f_j^2, $j = 1, \ldots, k$.

3. If there is a $j \in \{1, \dots, k\}$ such that $\left\langle f_j^2, x^k \right\rangle \neq 0$, reorder so that $\langle f_1^2, x^k \rangle \neq 0$. For $j = 2, ..., k$, define $f_j^3 = f_j^2 - b_j \cdot f_1^2$ with $b_j \in \mathbb{R}$ chosen so that $\left\langle f_j^3, x^k \right\rangle = 0$.

4. Iterate the above process for $x^{k+1}, \dots, x^{2k-2}$, where in each iteration the non-orthogonal function f_i^{i+1} is omitted going forward. This gives $k - i$ functions orthogonal to x^{k+i-1}. We obtain $f_1^2, f_2^3, \dots, f_k^{k+1}$ such that $\left\langle f_j^{j+1}, x^i \right\rangle = 0$ for $i \leq j + k - 2$.

5. Apply Gram Schmidt in order to $f_k^{k+1}, f_{k-1}^k, \dots, f_1^2$ to obtain $f_k, \dots, f_1$.

6. Define $a_j^I(x) = \frac{1}{\|f_j\|_2} f_j(x)$ for $j = 1, \dots, k$.

Informally, the procedure in step 3 "throws out" one function to achieve an extra degree of moment vanishing in the remaining functions. Although we will focus on moment vanishing, the same procedure could also be used to impose other conditions: continuity, differentiability, etc.

Alpert showed that this basis satisfies the same orthonormality, telescoping, and moment vanishing conditions as the Haar basis. We will delay the proof of this claim until the following section, where we prove the more general statement in Theorem 2.

Example 3. *Let $I = [0, 1)$ and $k = 3$. We want to construct the set $\{a_1^I, a_2^I, a_3^I\}$ where*

$$\left\langle a_j^I, x^m \right\rangle = 0, \quad j \in \{1, 2, 3\}, m \in \{0, 1, 2\}.$$

Because Alpert's algorithm uses Gram Schmidt, writing out explicit formulas for the functions becomes tedious very quickly. We will therefore not do so beyond the first step, but an example in full detail is given in section 3.3.2.

1. *We begin with*

$$f_1^1(x) = \mathbf{1}_{I_l} - \mathbf{1}_{I_r},$$
$$f_2^1(x) = x \cdot \mathbf{1}_{I_l} - x \cdot \mathbf{1}_{I_r},$$
$$f_3^1(x) = x^2 \cdot \mathbf{1}_{I_l} - x^2 \cdot \mathbf{1}_{I_r}.$$

2. *Apply the Gram Schmidt algorithm to the set $\{1, x, x^2, f_1^1, f_2^1, f_3^1\}$ in that order. Label the final three outputs f_1^2, f_2^2, f_3^2; each of these functions is piecewise quadratic on $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$, and each is orthogonal to 1, x, and x^2.*

3. *Define*

$$f_2^3 = f_2^2 - b_2 f_1^2 \quad \text{and} \quad f_3^3 = f_3^2 - b_3 f_1^2$$

with $b_2, b_3 \in \mathbb{R}$ *chosen so that* $\langle f_2^3, x^3 \rangle = 0$ *and* $\langle f_3^3, x^3 \rangle = 0$*. Both* f_2^3 *and* f_3^3 *are piecewise quadratic, and they are each orthogonal to* 1, x, x^2, *and* x^3.

4. *Define*

$$f_3^4 = f_3^3 - c_3 f_2^3$$

with $c_3 \in \mathbb{R}$ *chosen so that* $\langle f_3^4, x^4 \rangle = 0$*. So* f_3^4 *is piecewise quadratic, and is orthogonal to* 1, x, x^2, x^3, *and* x^4.

5. *We now have* $\{f_3^4, f_2^3, f_1^2\}$*; apply Gram Schmidt in that order and label the outputs* f_3, f_2, f_1*. Because* $\{f_3^4, f_2^3, f_1^2\}$ *are arranged from the most moment vanishing conditions to the least* f_2 *will possess the same moment vanishing properties as* f_2^3*, and likewise for* f_2 *and* f_2^3.

6. *The functions* f_1, f_2, *and* f_3 *are orthogonal and possess all the desired moment vanishing properties, so all that remains is to normalize:*

$$a_1^I(x) = \frac{1}{\|f_1\|_2} f_1(x), \ a_2^I(x) = \frac{1}{\|f_2\|_2} f_2(x), \ \text{and } a_3^I(x) = \frac{1}{\|f_3\|_2} f_3(x).$$

We have arrived at the orthonormal set $\{a_1^I, a_2^I, a_3^I\}$*, where all three functions are orthogonal to* 1, x, x^2*;* a_2^I *and* a_3^I *are additionally orthogonal to* x^3*, and* a_3^I *is also orthogonal to* x^4.

2.5 Weighted Alpert Wavelets

Sections 2.3 and 2.4 gave the basic constructions of Haar and Alpert wavelets on $\mathbb{R}$ in Lebesgue measure. These constructions can be generalized in three directions: by considering functions on $\mathbb{R}^n$ instead of $\mathbb{R}$, by replacing Lebesgue measure with an arbitrary measure, and by replacing the low-order polynomials in Alpert wavelets with arbitrary collections of L^2_{loc} functions. For the sake of expediency, we will give all three generalizations together and invite the interested reader to reverse-engineer the intermediate steps as special cases.

Definition 7 (Component space)**.** *Let* μ *be a locally finite positive Borel measure on* $\mathbb{R}^n$, $Q \in D$, *and* $U \subset L^2_{\text{loc}}(\mu)$. *The component space* $P_{Q,U}(\mu) = \text{Span}\{\mathbf{1}_Q \cdot p\}_{p \in U}$ *is the subspace of* $L^2_Q(\mu)$ *generated by the restrictions of* U *to* Q.

Let $\mathbb{E}^{\mu}_{Q,U}$ denote orthogonal projection onto $P_{Q,U}(\mu)$. With a slight abuse of notation we will allow a dyadic top $T \in \tau(D)$ in place of the dyadic cube Q, so $P_{T,U}(\mu)$ is the space of restrictions of U to T and $\mathbb{E}^{\mu}_{T,U}$ is the associated projection. Note that for many choices of μ, U, and T, the space $P_{T,U}(\mu)$ will be trivial since locally integrable functions may not have a finite integral when restricted to T.

Example 4. *Let* μ *be Lebesgue measure on* $\mathbb{R}$, $Q = [0,1)$, *and* $U = \{1, x\}$. *Then* $P_{Q,U}(\mu) = \text{Span}(\mathbf{1}_{[0,1)}, \mathbf{1}_{[0,1)}x)$. *Here* $\mathbf{1}_{[0,1)}$ *and* $\mathbf{1}_{[0,1)}x$ *form a basis for* $P_{Q,U}(\mu)$, *but this basis is not orthonormal. An example orthonormal basis is* $B = \{\mathbf{1}_{[0,1)}, \mathbf{1}_{[0,1)}(2\sqrt{3}x - \sqrt{3})\}$. *Then for any* $f \in L^2(\mu)$,

$$\mathbb{E}^{\mu}_{Q,U} = \langle f, \mathbf{1}_{[0,1)} \rangle \mathbf{1}_{[0,1)} + \left\langle f, \mathbf{1}_{[0,1)}(2\sqrt{3}x - \sqrt{3}) \right\rangle \cdot \mathbf{1}_{[0,1)}(2\sqrt{3}x - \sqrt{3}).$$

Example 5. *In the previous example the elements of* U *were a basis for* $P_{Q,U}(\mu)$, *but this is not always the case even when the elements of* U *are linearly independent in* $L^2(\mu)$. *To see how this can happen, let* $Q = [0,1)$ *and* $U = \{1, x\}$ *as before. Let* μ *assign Lebesgue measure for* $x < 0$, *but assign no mass for* $x \geq 0$ *except for a point mass at* $x = \frac{1}{2}$ *with* $\mu(\{\frac{1}{2}\}) = 1$. *Now the functions* $\mathbf{1}_{[0,1)}$, $\mathbf{1}_{[0,1)}x$ *are linearly dependent—in fact* $\mathbf{1}_{[0,1)} = 2 \cdot \mathbf{1}_{[0,1)}x$*—so* $P_{Q,U}(\mu)$ *is the one-dimensional space spanned by* $\mathbf{1}_{[0,1)}$.

Component spaces are of interest primarily because they are used to construct Alpert spaces, which are our primary object of study.

Definition 8 (Alpert space)**.** *Let* μ *be a locally finite positive Borel measure on* $\mathbb{R}^n$, $Q \in D$, *and* $U, V \subset L^2_{\text{loc}}(\mu)$. *The Alpert space* $L^2_{Q,U,V}(\mu)$ *is the subspace of functions in* $L^2_Q(\mu)$ *whose restrictions to each child* $Q' \in C(Q)$ *are in* $\mathbf{1}_{Q'}U$ *and which are orthogonal to each function in* V. *Namely:*

$$L^2_{Q,U,V}(\mu) = \left\{ f = \sum_{Q' \in C(Q)} \mathbf{1}_{Q'} p : \int_Q f(x) \cdot q(x)\, d\mu(x) = 0 \text{ for all } q(x) \in V \right\}$$

where each function p *is in* U.

Let $\triangle^{\mu}_{Q,U,V}$ denote orthogonal projection onto $L^2_{Q,U,V}(\mu)$. The term "moment vanishing conditions" is technically inaccurate when U is not a set of polynomials. We will however continue to use it to maintain consistency with earlier material, and hope that we have not offended the reader's moral sensibilities. Loosely speaking, Alpert spaces on a cube Q are the finite-dimensional vector spaces in which Alpert wavelets on Q live.

Example 6. *Let μ be Lebesgue measure on $\mathbb{R}$, $Q = [0, 1)$, and $U = V = \{1, x\}$. We have*

$$L^2_{Q,U,\varnothing}(\mu) = \mathrm{Span}(\mathbf{1}_{[0,\frac{1}{2})}, \mathbf{1}_{[0,\frac{1}{2})}x, \mathbf{1}_{[\frac{1}{2},1)}, \mathbf{1}_{[\frac{1}{2},1)}x),$$

and equivalently

$$L^2_{Q,U,\varnothing}(\mu) = P_{[0,\frac{1}{2}),U}(\mu) \oplus P_{[\frac{1}{2},1),U}(\mu).$$

In particular, the component space $P_{Q,U}(\mu) \subset L^2_{Q,U,\varnothing}(\mu)$. The Alpert space $L^2_{Q,U,V}(\mu)$ is the orthogonal complement of $P_{Q,U}(\mu)$ inside $L^2_{Q,U,\varnothing}(\mu)$, so it has dimension $4-2 = 2$. The calculations to produce an explicit orthonormal basis for $L^2_{Q,U,V}(\mu)$ are sufficiently cumbersome that we will not perform then here, but a complete example is presented in section 3.3.2.

Much of our work in chapter 3 broadly follows this example, decomposing an Alpert space into component spaces which can be evaluated individually. Rahm, Sawyer, and Wick [14] showed that Alpert's construction produces a basis for $L^2(\mu)$ with the same orthonormality, telescoping, and moment vanishing properties as in Lebesgue measure.

Theorem 2 (Weighted Alpert Bases)**.** *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, and $U \subseteq L^2_{\mathrm{loc}}(\mu)$ with $\mathbf{1} \in U$. Then*

$$\left\{\mathbb{E}^{\mu}_{T,U}\right\}_{T\in\tau(D)} \cup \left\{\triangle^{\mu}_{Q,U,U}\right\}_{Q\in D}$$

is a complete set of orthogonal projections in $L^2(\mu)$ and

$$f = \sum_{T\in\tau(D)} \mathbb{E}^{\mu}_{T,U} f + \sum_{Q\in D} \triangle^{\mu}_{Q,U,U} f, \qquad \text{for all } f \in L^2(\mu)$$

where convergence holds both in $L^2(\mu)$ and pointwise μ-almost everywhere. Moreover we have *orthogonality*

$$\left\langle \mathbb{E}^{\mu}_{T,U} f, \triangle^{\mu}_{Q,U,U} f \right\rangle = 0 = \left\langle \triangle^{\mu}_{R,U,U} f, \triangle^{\mu}_{Q,U,U} f \right\rangle \quad \textit{for all } Q \neq R \in D,$$

telescoping identities

$$\mathbf{1}_Q \sum_{P:\, Q \subsetneqq R \subseteq S} \triangle^{\mu}_{R,U,U} = \mathbb{E}^{\mu}_{Q,U} - \mathbf{1}_Q \mathbb{E}^{\mu}_{S,U} \textit{ for all } Q \subsetneqq S \in D,$$

and moment vanishing conditions

$$\int_{\mathbb{R}^n} \triangle^{\mu}_{Q,U,U} f(x) \cdot p(x)\, d\mu(x) = 0, \quad \textit{for all } Q \in D \textit{ and } p \in U.$$

Proof. Let $P, Q \in D$ with $P \neq Q$, and let $f \in L^2_{P,U,U}(\mu)$, $g \in L^2_{Q,U,U}(\mu)$ be a pair of Alpert functions. If P and Q are disjoint, then $\langle f, g\rangle = 0$ as f and g have disjoint support. Suppose instead that $P \subsetneqq Q$; then the restriction of g to P is equal to some function in $\mathbf{1}_P U$, and we get $\langle f, g\rangle = 0$ from the moment vanishing properties of $L^2_{P,U,U}(\mu)$. The same reasoning holds for $Q \subsetneqq P$, so we conclude that the Alpert spaces $L^2_{P,U,U}(\mu)$ and $L^2_{Q,U,U}(\mu)$ are orthogonal whenever $P \neq Q$.

Now fix a cube $P \in D$ and let $T \in \tau(D)$ be the top containing P. For $R \in D$ with $P \subsetneqq R$ we have

$$L^2_{P,\{1\},\{1\}}(\mu) \subseteq \overline{\mathrm{Span}} \left\{ \mathbf{1}_R U, \left\{ L^2_{Q,U,U}(\mu) \right\}_{Q \in D:\ Q \subseteq R \text{ and } l(P) \leq l(Q) \leq l(R)} \right\}.$$

Letting R tend to infinity, we conclude

$$L^2_{P,\{1\},\{1\}}(\mu) \subseteq \overline{\mathrm{Span}} \left\{ \mathbf{1}_T U, \left\{ L^2_{Q,U,U}(\mu) \right\}_{Q \in D:\ Q \subset T \text{ and } l(P) \leq l(Q)} \right\}. \tag{2.5.1}$$

The Haar spaces $L^2_{Q,\{1\},\{1\}}(\mu)$ and $\mathbf{1}_T$ form a direct sum decomposition of $L^2(\mu)$, i.e.

$$L^2(\mu) = \{\mathbf{1}_T\}_{T \in \tau(D)} \oplus \left(\bigoplus_{Q \in D} L^2_{Q,\{1\},\{1\}}(\mu) \right).$$

We can now apply 2.5.1 to each Alpert space in this decomposition and we have

$$L^2(\mu) = \{\mathbf{1}_T U\}_{T\in\tau(D)} \oplus \left(\bigoplus_{Q\in D} L^2_{Q,U,U}(\mu)\right).$$

Lastly we observe that the moment vanishing conditions follow from the definition of $L^2_{Q,U,U}(\mu)$, and the telecoping identities are an immediate consequence. This completes the proof. □

Remark 2. *It might seem strange that we have separated the sets U and V in our definition of $L^2_{Q,U,V}(\mu)$, only to immediately take $U = V$ for the main theorem. The separation will become useful when we want to consider applying extra moment vanishing conditions to some of the Alpert functions, i.e. requiring that some Alpert functions belong to $L^2_{Q,U,V}(\mu)$ for $U \subseteq V$.*

To emphasize a point that can get lost in the notation: the projections on a given dyadic top T in Theorem 2 are needed only when at least one function in U has a finite integral over T. In the particular case where U contains only polynomial functions, this condition is met if and only if T has finite μ-measure. This observation is due to Alexis, Sawyer, and Uriarte-Tuero [1].

2.6 Algebraic Geometry

Section 3.3 will investigate the special case of Alpert wavelet bases $L^2_{Q,U,U}(\mu)$ where U is taken to be the set of all monomials up to some fixed degree. The additional structure afforded by this choice allows for stronger conclusions than in the general case we explore in section 3.1. These results are found by applying tools from the field of algebraic geometry, which we summarize here.

For polynomials p and q, we will say p is a *multiple* of q—or equivalently that p is *divisible* by q—if there is some polynomial r such that $p = rq$. Denote by $\mathbb{R}[\mathbf{x}]$ the ring of real polynomials in n variables, where n will be understood from context.

Definition 9 (Algebraic Set)**.** *Let S be a set of polynomials in $\mathbb{R}[\mathbf{x}]$. The zero locus $Z(S) \subseteq \mathbb{R}^n$ is the set of common roots to every $p \in S$. An (affine) algebraic set V in $\mathbb{R}^n$ is any subset of $\mathbb{R}^n$ which is the zero locus of some S.*

In the event that S contains a single function p we will write $Z(p)$ as an abbreviation for $Z(\{p\})$.

Lemma 5. *Let S_1, S_2 be two finite sets of polynomials in $\mathbb{R}[\mathbf{x}]$. If $V_1 = Z(S_1)$ and $V_2 = Z(S_2)$ are two algebraic sets in $\mathbb{R}^n$, then $V_1 \cap V_2$ and $V_1 \cup V_2$ are also algebraic sets.*

Proof. $V_1 \cap V_2$ is exactly the set of points which are common roots to S_1 and S_2, so we have $V_1 \cap V_2 = Z(S_1 \cup S_2)$.

For $V_1 \cup V_2$, first define

$$S_1 \times S_2 := \{pq : p \in S_1, q \in S_2\}.$$

For any $x \in V_1$ we have $x \in Z(S_1 \times S_2)$ since every element in $S_1 \times S_2$ has x as a root. The same holds for $x \in V_2$ and so $V_1 \cup V_2 \subseteq Z(S_1 \times S_2)$.

Conversely, if $x \in Z(S_1 \times S_2)$ then x is a root of every polynomial pq with $p \in S_1$ and $q \in S_2$. Suppose that there is some $p_0 \in S_1$ such that x is not a root of p. Then since x is a root of $p_0 q$ for every $q \in S_2$, it follows that x is a root of every $q \in S_2$ and so $x \in V_2$. By an identical argument, if x is not a root of some $q_0 \in S_2$ then x must be in V_1. Therefore $Z(S_1 \times S_2) \subseteq V_1 \cup V_2$ and consequently $V_1 \cup V_2 = Z(S_1 \times S_2)$. □

This is sufficient for our purposes, but we will add a little expository background. The construction of $V_1 \cap V_2$ above clearly holds for arbitrary intersections, so this gives rise to a natural topology where the closed sets are precisely the algebraic sets. This topology is named the *Zariski topology*. An algebraic set V is called an *affine variety* if it cannot be expressed as the proper union of two algebraic subsets. Consequently any algebraic set can be expressed as a union of affine varieties.

Definition 10 (Monomial Order)**.** *A monomial order M on $\mathbb{R}[\mathbf{x}]$ is a well-ordering on the set of all (monic) monomials in n variables which respects multiplication. That is, for any monomials u, v, w we have $u < v$ implies $uw < vw$. A monomial order is graded if $x^\alpha < x^\beta$ implies $|\alpha| \leq |\beta|$ for any $\alpha, \beta \in \mathbb{N}^n$.*

For this work we will only use degree lexicographic order, in which monomials are ordered first by total degree, then by degree of x_1, then by degree of x_2, etc. This choice is for expositional simplicity; in practice there will often be some other choice of monomial order which achieves the same results but yields greater computational efficiency.

Definition 11 (Leading Term)**.** *Let M be a monomial order on $\mathbb{R}^n$. For a polynomial $p \in \mathbb{R}[\mathbf{x}]$, the leading term $LT(p)$ is the greatest monomial in p with respect to M. Similarly for any set S of polynomials, $LT(S)$ is the set of leading terms $\{LT(p)\}_{p\in S}$.*

Given an ideal $I \in \mathbb{R}[\mathbf{x}]$, the leading term ideal of I is the ideal $\langle LT(I)\rangle$ generated by the leading terms of I. Note that the set of leading terms $LT(I)$ is *not* itself an ideal, since the sum of two distinct monomials is generally not a monomial.

Definition 12 (Gröbner Basis)**.** *Let M be a monomial order on $\mathbb{R}^n$ and let I be an ideal in $\mathbb{R}[\mathbf{x}]$. A Gröbner basis G for I is a generating set for I such that for any polynomial $p \in I$, $LT(p)$ is divisible by $LT(q)$ for some $q \in G$. A Gröbner basis is called reduced if no monomial in any $p \in G$ is in $LT(G \setminus \{p\})$ and every $p \in G$ is monic.*

An equivalent definition is: a Gröbner basis G is a generating set for I such that $\langle LT(I)\rangle = \langle LT(G)\rangle$. While Gröbner bases depend on the choice of monomial order and in general are not unique, they are guaranteed to exist for polynomial rings over a field in finitely many variables. Moreover, for a given monomial order there is a unique reduced Gröbner basis.

Gröbner bases are efficiently computable for any choice of monomial basis, although the details of such computations are not important for our results. For the interested reader we provide one such algorithm in appendix B.

Definition 13 (Hilbert Dimension)**.** *Let I be an ideal in $\mathbb{R}[\mathbf{x}]$. The Hilbert dimension of I is the maximal size of a subset S of variables in $\mathbf{x}$ such that no leading monomial in I can be expressed entirely using variables in S.*

Observe that Hilbert dimension can be easily computed using a Gröbner basis; given an ideal I with Gröbner basis G, the set of leading terms $LT(I)$ is equal to $LT(G)$. So the size of a maximal set of variables which produces none of the elements in $LT(G)$ gives the Hilbert dimension of I.

Chapter 3

Structure of Weighted Alpert Wavelets

3.1 Dimensions of Alpert Spaces

In this section we study the structure of individual Alpert spaces $L^2_{Q,U,V}(\mu)$. Our goal is to concretely describe a method for finding the dimensions of such spaces.

Recall the informal definition of $L^2_{Q,U,V}(\mu)$: this is the space of functions which are piecewise sums of functions in $\mathbf{1}_{Q'}U$, with Q' varying over the children of Q, subject to moment vanishing conditions given by $\mathbf{1}_Q V$. This leads to a natural first guess; perhaps the dimension of $L^2_{Q,U,V}(\mu)$ is simply the sum of the dimensions of each $\mathrm{Span}(\mathbf{1}_{Q'}U)$ minus the dimension of $\mathrm{Span}(\mathbf{1}_Q V)$. This calculation does give the correct dimension for the Lebesgue Haar and Alpert wavelet bases described in sections 2.3 and 2.4. Alas, the pattern does not hold in general.

Example 7. *Take μ to be Lebesgue measure on $\mathbb{R}$, and D to be a dyadic grid containing the interval $I = [-1, 1)$. If we take $U = V_0 = \{1\}$ then $L^2_{I,U,V_0}(\mu)$ is the usual one-dimensional Haar space spanned by $\mathbf{1}_{I_l} - \mathbf{1}_{I_r}$. If we now expand the set of moment vanishing conditions to $V_1 = \{1, x\}$, we see that $L^2_{I,U,V_0}(\mu)$ contains no non-trivial functions orthogonal to x and so $L^2_{I,U,V_1}(\mu)$ has dimension zero. This is the behaviour our naive guess expected; adding an extra moment vanishing condition reduced the dimension of the Alpert space by one.*

To break the pattern, instead take $V_2 = \{1, x^2\}$. Now we see that the Haar functions in

$L^2_{I,U,V_0}(\mu)$ are already orthogonal to x^2—indeed they are orthogonal to every even function on $\mathbb{R}$. Therefore $L^2_{I,U,V_0}(\mu)$ and $L^2_{I,U,V_2}(\mu)$ are the same space and have the same dimension, despite the additional moment vanishing condition.

This shows that considering the various parameters in isolation is insufficient to produce an answer in the general case. Instead we consider the effect of adding a single new moment vanishing condition to an Alpert space.

Theorem 3. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, $Q \in D$, $U, V \subseteq L^2_{\text{loc}}(\mu)$ be finite sets, and $f \in L^2_{\text{loc}}(\mu)$ be a function such that $f \notin V$. Then either* $\dim L^2_{Q,U,V\cup\{f\}}(\mu) = \dim L^2_{Q,U,V}(\mu)$*, if f is orthogonal to all of $L^2_{Q,U,V}(\mu)$, or* $\dim L^2_{Q,U,V\cup\{f\}}(\mu) = \dim L^2_{Q,U,V}(\mu) - 1$.

Proof. If f is orthogonal to all of $L^2_{Q,U,V}(\mu)$ then $L^2_{Q,U,V\cup\{f\}}(\mu)$ and $L^2_{Q,U,V}(\mu)$ are the same space and trivially have the same dimension. Suppose instead that f is not orthogonal to all of $L^2_{Q,U,V}(\mu)$. Let $\{b_1, ..., b_k\}$ be a basis for $L^2_{Q,U,V}(\mu)$ and suppose without loss of generality that f is not orthogonal to b_1.

Now for $i = 2, ..., k$, define the constant $c_i \in \mathbb{R}$ as

$$c_i = -\frac{\int_Q b_i(x)f(x)\,d\mu(x)}{\int_Q b_1(x)f(x)\,d\mu(x)}.$$

This gives

$$\langle b_i + c_i b_1, f\rangle = \int_Q b_i(x)f(x)\,d\mu(x) - \frac{\int_Q b_i(x)f(x)\,d\mu(x)}{\int_Q b_1(x)f(x)\,d\mu(x)} \int_Q b_1(x)f(x)\,d\mu(x) = 0$$

so $b_i + c_i b_1$ is orthogonal to f. Now suppose that $\{b_i + c_i b_1\}_{i=2,...,k}$ contains a linear dependence. Then we would have

$$\sum_{i=2}^{k} a_i(b_i + c_i b_1) = 0$$

for some constants $a_i \in \mathbb{R}$, $i = 2, ...k$ not all zero. Rearranging gives

$$\left(\sum_{i=2}^{k} a_i c_i\right) b_1 + \sum_{i=2}^{k} a_i b_i = 0$$

which is impossible as $\{b_1, ..., b_k\}$ is a basis for $L^2_{Q,U,V}(\mu)$. Consequently $\{b_i + c_i b_1\}_{i=2,...,k}$ is linearly independent and forms a basis for $L^2_{Q,U,V\cup\{f\}}(\mu)$, and we conclude that $\dim L^2_{Q,U,V\cup\{f\}}(\mu) = \dim L^2_{Q,U,V}(\mu) - 1$ □

The key insight here is that $L^2_{Q,U,\varnothing}(\mu)$ is a finite-dimensional vector space, and that any $L^2_{Q,U,V}(\mu)$ is a subspace of $L^2_{Q,U,\varnothing}(\mu)$. In particular, suppose that $V = \{v_1, ...v_k\}$ is a finite set of functions none of which are orthogonal to $L^2_{Q,U,\varnothing}(\mu)$. Then by Theorem 3, each $L^2_{Q,U,\{v_i\}}(\mu)$ is a hyperplane inside $L^2_{Q,U,\varnothing}(\mu)$, and

$$L^2_{Q,U,V}(\mu) = \bigcap_{i=1,\dots,k} L^2_{Q,U,\{v_i\}}(\mu).$$

We know from elementary linear algebra that k hyperplanes can intersect in a subspace which has dimension anywhere from 1 to k less than the ambient space, and in light of this cases like Example 7 are unsurprising.

Corollary 1. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, $Q \in D$, and $U, V \subseteq L^2_{\mathrm{loc}}(\mu)$ be finite sets. Then*

$$\sum_{Q'\in C(Q)} \dim P_{Q',U}(\mu) - \dim P_{Q,V}(\mu) \leq \dim L^2_{Q,U,V}(\mu) \leq \sum_{Q'\in C(Q)} \dim P_{Q',U}(\mu).$$

Proof. Let $V_0 \subseteq V$ be a maximal subset such that $\mathbf{1}_Q V_0$ is linearly independent. Suppose that $f \in L^2_{\mathrm{loc}}(\mu)$ is a function which can be expressed as a linear combination of the elements in V_0. Since $L^2_{Q,U,V_0}(\mu)$ is orthogonal to all of V_0, it is also orthogonal to f. Then by applying the first conclusion of Theorem 3 to each $f \in V \setminus V_0$, we conclude that $L^2_{Q,U,V}(\mu)$ and $L^2_{Q,U,V_0}(\mu)$ are the same space.

Now since $\dim P_{Q,V}(\mu) = \dim P_{Q,V_0}(\mu) = \#V_0$, applying Theorem 3 to each element in V_0 gives

$$\dim L^2_{Q,U,\varnothing}(\mu) - \#V_0 \leq \dim L^2_{Q,U,V_0}(\mu) \leq \dim L^2_{Q,U,\varnothing}(\mu)$$

and consequently

$$\sum_{Q'\in C(Q)} \dim P_{Q',U}(\mu) - \dim P_{Q,V}(\mu) \leq \dim L^2_{Q,U,V}(\mu) \leq \sum_{Q'\in C(Q)} \dim P_{Q',U}(\mu)$$

as desired. □

Theorem 3 is, in a sense, a double-edged sword. On the one hand, it does not give a simple closed form for finding $\dim L^2_{Q,U,V}(\mu)$. Without investigating the underlying geometry in a given case, we cannot conclude more about an Alpert space's dimension than the inequality given in Corollary 1.

On the other hand, when it comes to construction of a particular Alpert basis, this is not a serious problem. The algorithm for constructing Alpert bases—given in Section 2.4—already requires finding a non-orthogonal basis function for each orthogonality condition to be introduced. Any "freebies" among the set of moment vanishing conditions will be discovered in the course of performing the algorithm, and Theorem 3 guarantees that there are no other kinds of bad behaviour to be concerned about.

Lastly, we emphasize that these considerations are only a concern when the additional moment vanishing properties allowed by Alpert bases are of interest; this is equivalent to asking that a basis for $L^2_{Q,U,U}(\mu)$ additionally have some elements belonging to $L^2_{Q,U,V}(\mu)$ for some $V \supseteq U$. If any basis for $L^2_{Q,U,U}(\mu)$ will suffice, the situation is much simpler.

Corollary 2. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, $Q \in D$, and $U, V \subseteq L^2_{\text{loc}}(\mu)$ be finite sets such that $V \subseteq U$. Then*

$$\dim L^2_{Q,U,V}(\mu) = \sum_{Q' \in C(Q)} \dim P_{Q',U}(\mu) - \dim P_{Q,V}(\mu).$$

Proof. Since $V \subseteq U$, we have $P_{Q,V} \subseteq L^2_{Q,U,\varnothing}(\mu)$. $P_{Q,V}$ is therefore the orthogonal complement of $L^2_{Q,U,V}(\mu)$ inside $L^2_{Q,U,\varnothing}(\mu)$ as finite-dimensional vector spaces, and the result follows immediately. □

So the problem of finding dimensions and bases for Alpert spaces $L^2_{Q,U,U}(\mu)$ reduces to the equivalent problem for component spaces $P_{Q,U}(\mu)$. A maximal set of linearly independent functions in U gives a basis for $P_{Q,U}(\mu)$, so given Q, U, and μ it suffices to find such a set. A brute force algorithm could proceed as follows: begin with the entire set U and iteratively apply the Gram Schmidt process to identify and eliminate dependent functions until a basis is found. This is rather inefficient, so in section 3.3 we will give a more sophisticated algorithm when working with polynomial Alpert bases.

3.2 Variable Alpert Bases

The standard construction of an Alpert basis chooses some set of orthogonality conditions to be applied to the basis functions on each dyadic cube. In this section we show that this can be relaxed: each cube Q can be given a differing set of orthogonality conditions, and with a small modification the resulting functions still form an orthonormal basis for $L^2(\mu)$.

Let D be a dyadic grid on $\mathbb{R}^n$ and $Q \in D$. We define the *tower* $\Gamma(Q)$ as the set of all cubes in D which contain Q. The *top* T of a tower $\Gamma(Q)$ is defined as the countable union of all cubes in the tower. As an immediate consequence of the nesting property of towers, any two towers will either have the same top or two disjoint tops. This gives an equivalence relation on the towers in D, where Γ_1 and Γ_2 are equivalent if their tops coincide. Let $\tau(D)$ denote the set of unique tops arising from any choice of representatives from each equivalence class.

Alexis, Sawyer, and Uriarte-Tuero in [1] observed that, besides bases for each $L^2_{Q,U,U}$, a standard Alpert basis may also require the restrictions to some dyadic tops of functions in U. Specifically, for every $f \in U$ and $T \in \tau(D)$, if $\mathbf{1}_T f$ has finite $L^2(\mu)$-norm then it must be included in the Alpert basis. This remains true for our generalization.

Let $\mathbb{E}^\mu_{Q,U}$ denote orthogonal projection onto $P_{Q,U}(\mu)$. With a slight abuse of notation we will allow a dyadic top $T \in \tau(D)$ in place of the dyadic cube Q, so $P_{T,U}(\mu)$ is the space of restrictions of U to T and $\mathbb{E}^\mu_{T,U}$ is the associated projection. Let $\triangle^\mu_{Q,U,V}$ similarly denote orthogonal projection onto $L^2_{Q,U,V}(\mu)$.

For sets of functions $U \subseteq V$ we will also need to consider the orthogonal complement of $P_{Q,U}(\mu)$ inside $P_{Q,V}(\mu)$. As our notation is already somewhat cumbersome, we will write $P_{Q,V}(\mu) \ominus P_{Q,U}(\mu)$ to denote such a subspace and $\mathbb{E}^\mu_{Q,V} - \mathbb{E}^\mu_{Q,U}$ to denote the corresponding projection. Lastly, recall that $P(Q)$ denotes the parent of Q.

Theorem 4. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, and $\{U_Q\}_{Q\in D}$ be a collection of finite sets in $L^2_{\text{loc}}(\mu)$ and having the following properties:*

1. *$1 \in U_Q$ for every $Q \in D$.*

2. *$U_Q \subseteq U_{Q'}$ for every $Q \in D$ and every $Q' \in C(Q)$.*

For each dyadic top $T \in \tau(D)$, let $U_T = \cap_{Q \subset T} U_Q$. Then

$$\left\{\triangle^{\mu}_{Q,U_Q,U_Q}\right\}_{Q\in D} \cup \left\{\mathbb{E}^{\mu}_{Q,U_Q} - \mathbb{E}^{\mu}_{Q,U_{P(Q)}}\right\}_{Q\in D} \cup \left\{\mathbb{E}^{\mu}_{T,U_T}\right\}_{T\in\tau(D)}$$

is a complete set of orthogonal projections in $L^2(\mu)$ and

$$f = \sum_{Q\in D} \triangle^{\mu}_{Q,U_Q,U_Q} f + \sum_{Q\in D} \left(\mathbb{E}^{\mu}_{Q,U_Q} - \mathbb{E}^{\mu}_{Q,U_{P(Q)}}\right) f + \sum_{T\in\tau(D)} \mathbb{E}^{\mu}_{T,U_T} f, \qquad f \in L^2(\mu)$$

where convergence holds both in $L^2(\mu)$ and pointwise μ-almost everywhere. Moreover we have telescoping identities,

$$\mathbf{1}_Q \sum_{P:Q \subsetneqq P \subseteq R} \triangle^{\mu}_{P,U_P,U_P} = \mathbb{E}^{\mu}_{Q,U_Q} - \mathbf{1}_Q \mathbb{E}^{\mu}_{R,U_R} \text{ for all } Q \subsetneqq R \in D \text{ with } U_Q = U_R,$$

and moment vanishing conditions

$$\int_{\mathbb{R}^n} \triangle^{\mu}_{Q,U_Q,U_Q} f(x) p(x)\, d\mu(x) = 0, \quad \text{for all } Q \in D, p \in U_Q.$$

Here the $\triangle^{\mu}_{Q,U_Q,U_Q}$ are the usual Alpert projections, and the $\mathbb{E}^{\mu}_{T,U_T}$ are the aforementioned projections on the tops of D. The $\mathbb{E}^{\mu}_{Q,U_Q} - \mathbb{E}^{\mu}_{Q,U_{P(Q)}}$ are the necessary addition which allows the basis to retain completeness while varying U_Q, which we prove now. Besides this consideration, we otherwise follow the strategy used by Rahm, Sawyer, and Wick in [14, Theorem 1] to show the equavialent result for polynomial Alpert bases.

Proof. Fix $Q \in D$. By construction we have

$$L^2_{Q,U_Q,U_Q}(\mu) \oplus P_{Q,U_Q}(\mu) = L^2_{Q,U_Q,\varnothing}(\mu). \tag{3.2.1}$$

Applying this to each child $Q' \in C(Q)$ we have

$$\bigoplus_{Q'\in C(Q)} P_{Q',U_{Q'}}(\mu) \oplus \bigoplus_{Q'\in C(Q)} L^2_{Q',U_{Q'},U_{Q'}}(\mu) = \bigoplus_{Q'\in C(Q)} L^2_{Q',U_{Q'},\varnothing}(\mu).$$

The leftmost sum in this expression can be rewritten as

$$\bigoplus_{Q' \in C(Q)} P_{Q',U_{Q'}}(\mu) = L^2_{Q,U_Q,U_Q}(\mu) \oplus P_{Q,U_Q}(\mu) \oplus \bigoplus_{Q' \in C(Q)} \Big(P_{Q',U_{Q'}} \ominus P_{Q',U_Q}(\mu) \Big) .$$

The nesting property $U_Q \subseteq U_{Q'}$ guarantees that this construction is valid.

Next consider the tower $\Gamma(Q)$ and choose some $S \in \Gamma(Q)$. Let X be the set of all cubes R contained in S and with side lengths $l(Q) < l(R) \leq l(S)$. Applying the above argument iteratively to every $P \in X$ yields

$$\bigoplus_{\substack{Q' \subset S \\ l(Q')=l(Q)}} L^2_{Q',U_{Q'},\varnothing}(\mu) = \bigoplus_{R \in X} L^2_{R,U_R,U_R}(\mu) \oplus P_{S,U_S}(\mu) \oplus \bigoplus_{R \in X} \left(P_{R,U_R} \ominus P_{R,U_{P(R)}}(\mu) \right) .$$

In particular, this sum contains the Haar space $L^2_{Q,\{1\},\{1\}}(\mu)$ since $1 \in U_Q$.

Now take the limit of this construction as S tends to infinity. Let T be the top of $\Gamma(Q)$; in the limit, $P_{S,U_S}(\mu)$ becomes $P_{T,U_T}(\mu)$. Thus for any $Q \in D$ we have

$$L^2_{Q,\{1\},\{1\}}(\mu) \subseteq \bigoplus_{Q \in D} L^2_{Q,U_Q,U_Q}(\mu) \oplus \bigoplus_{T \in \tau(D)} P_{T,U_T}(\mu) \oplus \bigoplus_{Q \in D} \left(P_{Q,U_Q} \ominus P_{Q,U_{P(Q)}}(\mu) \right) .$$

We know that the Haar spaces $L^2_{Q,\{1\},\{1\}}(\mu)$, $Q \in D$, together with projections on the tops $P_{T,\{1\}}(\mu)$, $T \in \tau(D)$, form a direct sum decomposition of $L^2(\mu)$. We also have $P_{T,\{1\}}(\mu) \subseteq P_{T,U_T}(\mu)$ since $1 \in U_Q$ for every $Q \in D$, so we conclude

$$L^2(\mu) = \bigoplus_{Q \in D} L^2_{Q,U_Q,U_Q}(\mu) \oplus \bigoplus_{T \in \tau(D)} P_{T,U_T}(\mu) \oplus \bigoplus_{Q \in D} \left(P_{Q,U_Q} \ominus P_{Q,U_{P(Q)}}(\mu) \right) .$$

The moment vanishing conditions are satisfied by construction: the nesting property $U_Q \subseteq U_{Q'}$ ensures that any $\triangle^{\mu}_{Q,U_Q,U_Q} f$ is orthogonal to every $p \in U_R$ for any $R \in D$ containing Q. Lastly the telescoping property follows by chaining together instances of (1), exactly as in the standard Alpert construction. □

This construction partially alleviates one of the drawbacks of Alpert bases, namely that extra orthogonality is achieved only by making the basis much larger than its Haar counterpart. This might be of interest in applications where the additional orthogonality is only needed locally, or only beyond a particular level of resolution. We end with a pair of remarks concerning the structure of the resulting basis.

Remark 3. *The telescoping identity in Theorem 4 is weaker than in a standard Alpert basis where all U_Q are equal. However since we still have $U_R \subseteq U_Q$ for $Q \subseteq R$, multiple instances of this weaker identity can be chained together to produce an expression that is still "good". At each level in the chain one subtracts projections only onto the functions which are in $U_{P'}$ but not in U_P, for $P' \in C(P)$.*

Remark 4. *The projections at the tops can be interpreted as a special case of the projections $\mathbb{E}^{\mu}_{Q,U_Q} - \mathbb{E}^{\mu}_{Q,U_{P(Q)}}$. If we think of a top T as being a kind of dyadic cube and define $U_{P(T)} = \varnothing$, as tops do not have parents, then $\mathbb{E}^{\mu}_{Q,U_{P(T)}}$ is trivial and we recover the usual projection on T. In this sense our generalization only extends an existing complexity in Alpert bases, rather than introducing a new one.*

Figure 3.2.1 on the following page summarizates the relations between the various types of Haar and Alpert wavelet basis we have seen in this work. For each type of basis we show the corresponding decomposition of L^2 into Alpert spaces, as well as references for each.

3.3 Structure Theorem for Polynomial Alpert Bases

In section 3.1, we saw that the dimension of an Alpert space $L^2_{Q,U,V}(\mu)$ depended non-trivially on the underlying geometry of both the measure μ and the set of functions U, V in question. We now turn our attention to the specific case of polynomial Alpert bases; the additional structure this affords will allow us to draw more precise conclusions. We consider spaces $L^2_{Q,U,U}(\mu)$ where U is taken to be the set of all monomials in n variables up to some fixed degree. Since we will use such sets heavily in this section it is useful to have the following notation.

Definition 14. *Given $k, n \in \mathbb{N}$, define F^n_k to be the set of all monomials in n variables with degree less than k.*

Lemma 6. *Given $k, n \in \mathbb{N}$, the number of monomials in n variables with degree less than k is $\#F^n_k = \binom{n+k-1}{n}$.*

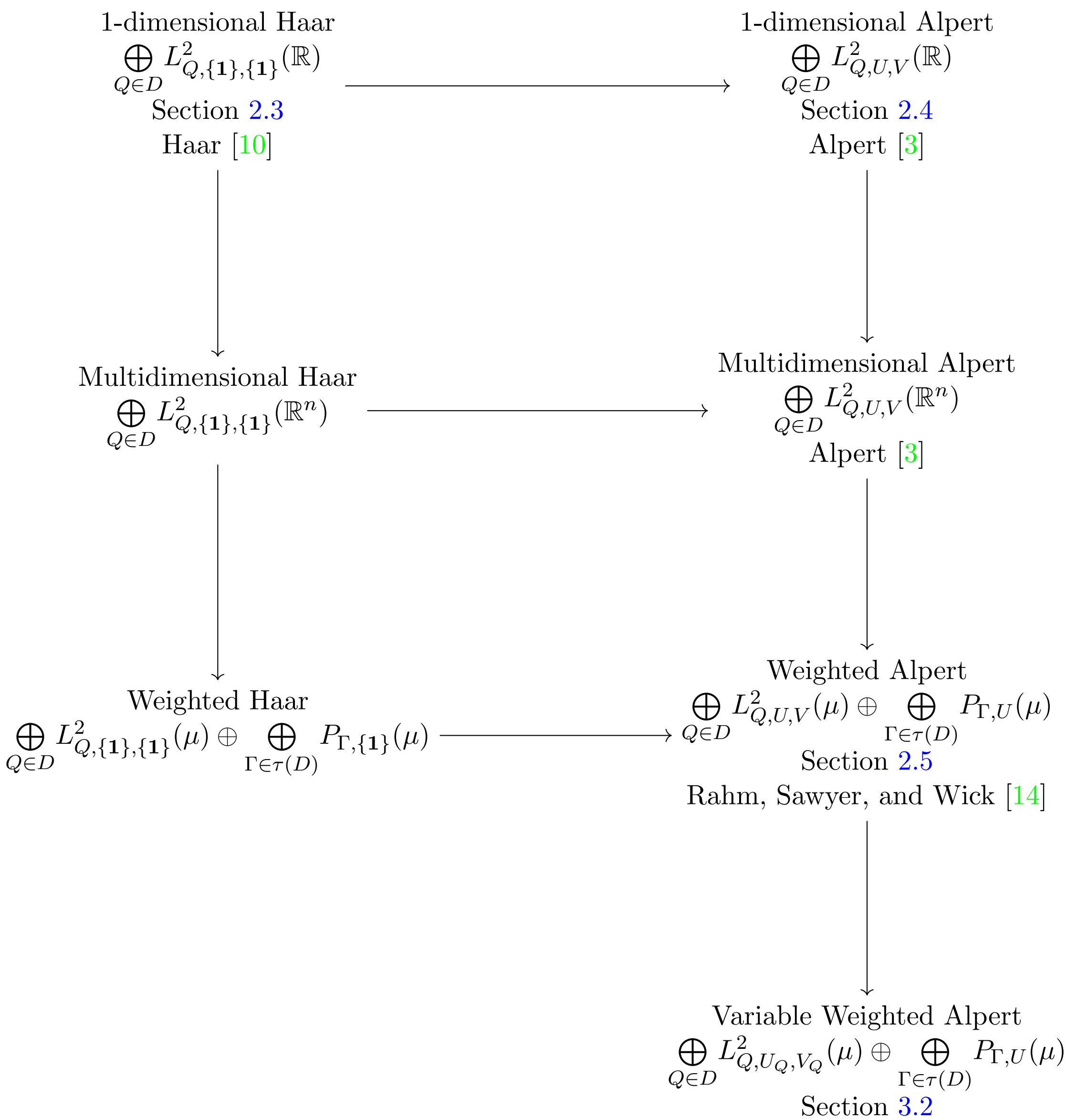

Figure 3.2.1: Relations among classes of wavelet basis

Proof. Consider a string containing $k-1$ copies of the symbol $\star$ and n copies of the symbol $|$, arranged in some arbitrary order. Such a string is a graphical representation of the partitioning of the integer $k-1$ into a sum of $n+1$ non-negative integers; the elements of the sum are the numbers of consecutive $\star$'s. Alternatively, the same string can also be interpreted as partitioning an integer i, $0 \leq i \leq k-1$ into n non-negative integers by discarding the final $|$ and any following $\star$'s.

Let x^α be a monomial in n variables with degree less than k. The exponent α is precisely an n-tuple of non-negative integers whose sum is at most $k-1$ (and at least 0). Therefore we have a one-to-one correspondence between the set of monomials in n variables with degree less than k and the set of partitions of all integers i, $0 \leq i \leq k-1$. From above, these partitions also have a one-to-one correspondence with the set of strings containing $k-1$ $\star$'s and n $|$'s, and there are $\binom{n+k-1}{n}$ such strings. We conclude that $\#F_k^n = \binom{n+k-1}{n}$. □

Example 8. *To illustrate the above counting argument, let $n = 4$, $k = 9$, and consider the monomial $x^\alpha = x_1^3 \cdot x_2 \cdot x_4^2$. The analogous string is:*

$$\star\star\star|\star||\star\star|\star\star.$$

The two consecutive $|$'s indicate that x_3 has an exponent of 0, and the final two $\star$'s are discarded so that we have a monomial of degree 6 (from the maximum allowed degree of 8). The number of strings containing 8 $\star$'s and 4 $|$'s is

$$\binom{4+8}{4} = \binom{12}{4} = 495$$

so there are 495 unique monomials in 4 variables with degree less than 9.

In [14], Rahm, Sawyer, and Wick observed that polynomial Alpert spaces over certain measures had lower dimension than the corresponding spaces would have over Lebesgue measure—indeed, it was that observation which first motivated this thesis. From Corollary 2 we have

$$\dim L^2_{Q,F_k^n,F_k^n}(\mu) = \sum_{Q' \in C(Q)} \dim P_{Q',F_k^n}(\mu) - \dim P_{Q,F_k^n}(\mu).$$

This behaviour is therefore explained by observing that, unlike in Lebesgue measure, the monomials in F_k^n are not guaranteed to be linearly independent and consequently that $\dim P_{Q,F_k^n}(\mu)$ might be less than $\#F_k^n$.

This reduces the question of finding $\dim L^2_{Q,F_k^n,F_k^n}(\mu)$ to the question of finding $\dim P_{Q,F_k^n}(\mu)$ for arbitrary Q. It is not immediately obvious that this can be done easily; even if some subset of F_k^n is linearly independent in $L^2(\mu)$, it might be linearly dependent in $L^2_Q(\mu)$ for some Q. As a simple example, all monomials (and indeed all functions) are linearly dependent in $L^2_Q(\mu)$ if μ assigns weight to only a single point mass inside Q.

Definition 15. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, and $Q \in D$. Define A_Q to be the intersection of all algebraic sets A such that $\mu(Q \setminus A) = 0$*

Definition 16. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, and $Q \in D$. Define I_Q to be the ideal of all polynomials in $\mathbb{R}[\mathbf{x}]$ which vanish on A_Q.*

Informally, A_Q and I_Q give a canonical representation of all the linear dependences among F_k^n in $L^2_Q(\mu)$. Computing a generating set for I_Q given A_Q is a difficult problem, in the sense that there is no general algorithm for producing such sets and individual cases must be attacked heuristically. However, provided that we can find a generating set for I_Q, we can then leverage a Gröbner basis to compute a basis for $P_{Q,F_k^n}(\mu)$.

Theorem 5. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, $Q \in D$, and $k \in \mathbb{N}$. Also let M be a graded monomial order and G be the reduced Gröbner basis for I_Q. Define $\operatorname{dep}_k(G)$ to be the set of all monomials $u \in F_k^n$ which are divisible by some monomial $v \in LT(G)$, and define $\operatorname{ind}_k(G)$ to be the complement of $\operatorname{dep}_k(G)$ in F_k^n. Then $\operatorname{ind}_k(G)$ is a basis for $P_{Q,F_k^n}(\mu)$, and consquently $\dim P_{Q,F_k^n}(\mu) = \#F_k^n - \#\operatorname{dep}_k(G)$.*

Proof. Suppose $\operatorname{ind}_k(G)$ contains a linear dependence in $L^2_Q(\mu)$, given by some polynomial $p(x) = 0$. Then $p \in I_Q$, so $LT(p) \in LT(G)$ and $LT(p) \notin \operatorname{ind}_k(G)$. This is a contradiction, so $\operatorname{ind}_k(G)$ is linearly independent in $L^2_Q(\mu)$ and it remains to show that $\operatorname{ind}_k(G)$ is maximal.

Let $u_0 \in \operatorname{dep}_k(G)$ and consider the set $T := \{u_0\} \cup \operatorname{ind}_k(G)$. Since $u_0 \in LT(G)$, there is some monomial $v_0 \in F_k^n$ and some polynomial $p_0 \in G$ such that $u_0 = LT(v_0 p_0)$. Then since the monomial order M is graded and $LT(v_0 p_0)$ has degree less than k, all monomials in $v_0 p_0$ must have degree less than k and so $v_0 p_0$ is a linear dependence in F_k^n. It now suffices to show that $v_0 p_0$ can be expressed using only monomials in T.

Let $u_1 \notin T$ be a non-leading monomial occurring in $v_0 p_0$. Since $u_1 \in \mathrm{dep}_k(G)$ there is some monomial $v_1 \in F_k^n$ and some polynomial $p_1 \in G$ such that $u_1 = LT(v_1 p_1)$. The non-leading monomials in $v_1 p_1$ all have order less than u_1, so we can replace u_1 in $v_0 p_0$ with lower order monomials. Since $\mathrm{dep}_k(G)$ is finite, iterating this process a sufficient number of times yields a representation for $v_0 p_0$ using only monomials in T. Consequently T is linearly dependent in $L^2_Q(\mu)$ for any choice of $u_0 \in \mathrm{dep}_k(G)$, so we conclude that $\mathrm{ind}_k(G)$ must be a basis for $P_{Q,F_k^n}(\mu)$. □

We can now justify our restriction to the case of F_k^n, rather than arbitrary collections of monomials: these sets have the special property that any monomial not in F_k^n is greater than any monomial in F_k^n with respect to the graded monomial order M. If we take an arbitrary set of monomials U which does not have this property, then even if an element of U is divisible by some element of $LT(G)$ the corresponding linear dependence may involve monomials which are not in U. Since a monomial order must respect multiplication, most sets U will not allow a choice of monomial order which avoids this problem.

Fortunately, this result is sufficient to construct the desired basis for $L^2_{Q,F_k^n,F_k^n}(\mu)$. Theorem 5 finds a monomial basis for $P_{Q,F_k^n}(\mu)$ and every $P_{Q',F_k^n}(\mu)$ with $Q' \in C(Q)$. Then Alpert's projection technique from section 2.4 imposes the necessary orthogonality on the basis elements. As a final step, an orthonormal basis can be achieved by applying Gram-Schmidt in reverse order, beginning with the basis function with the most additional moment vanishing and ending with the least.

3.3.1 Suuplemental Results

In addition to our main result from the previous section, we have collected a scattering of more minor observations regarding polynomial Alpert bases. We present them here.

Lemma 7. *Let I be an ideal in $\mathbb{R}[\mathbf{x}]$ with Hilbert dimension d. Then there exists $k \in \mathbb{N}$ such that $x^\alpha \in LT(I)$ for all $\alpha \in \mathbb{N}^n$ with more than d entries greater than k.*

Proof. Suppose toward a contradiction that there is no such k. Then for every $k \in \mathbb{N}$ we can associate an $x^{\alpha_k} \notin LT(I)$ where at least $d+1$ entries in α_k are greater than k. This yields the sequence of monomials $\{x^{\alpha_k}\}_{k \in \mathbb{N}}$. Since there are only finitely many variables to choose from, there must be a set $S = \{x_{i_1}, x_{i_2}, \ldots, x_{i_{d+1}}\}$ of $d+1$ distinct variables such that for any $l \in \mathbb{N}$ the product $x_{i_1}^l \cdot x_{i_2}^l \cdots x_{i_{d+1}}^l$ divides some monomial in $\{x^{\alpha_k}\}_{k\in\mathbb{N}}$. Since no x^{α_k} is in $LT(I)$, we conclude that $x_{i_1}^l \cdot x_{i_2}^l \cdots x_{i_{d+1}}^l \notin LT(I)$ for all $l \in \mathbb{N}$.

Now we see that any product of the variables in S divides $x_{i_1}^l \cdot x_{i_2}^l \cdots x_{i_{d+1}}^l$ for some sufficiently large choice of l, and consequently any product of these variables produces a monomial not in $LT(I)$. Therefore S shows that I must have Hilbert dimension at least $d+1$. This completes the contradiction, so such a k must exist. □

Building on this idea, we have the following intuition: by theorem 5, the leading terms of a Gröbner basis for I_Q determine the linearly dependent monomials that need to be excluded to form a basis for $P_{Q,F_k^n}(\mu)$. After making all of these exclusions, what remains should grow like a Lebesgue polynomial function space. While this doesn't allow us to compute $\dim P_{Q,F_k^n}(\mu)$ directly, it does characterize the growth rate of this dimension as k increases. Here we use big O notation in its usual meaning: $f(x) = O(g(x))$ if for some constant $C > 0$ and some x_0 we have $|f(x)| \leq C \cdot g(x)$ for all $x > x_0$.

Theorem 6. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, D be a dyadic grid, $Q \in D$, and $k \in \mathbb{N}$. Suppose that I_Q has Hilbert dimension d. Then $\dim P_{Q,F_k^n}(\mu) = O(k^d)$.*

Proof. If $d = n$ we are immediately done, so suppose that $d < n$. For any set S' of $d+1$ variables, consider all monomials that can be expressed using only variables in S'. At least one such monomial must be a leading monomial within I_Q, otherwise this set would give I_Q a Hilbert dimension of at least $d+1$. Thus we cannot have an independent monomial containing arbitrarily high powers of all $d+1$ variables in S'. For any choice of variable to have a bounded power, there are $O(k^d)$ such monomials. As there are only finitely many variables to choose, the total number of independent monomials using only variables in S' is at most $O(k^d)$.

Now we generalize this argument to any set of more than d variables. By lemma 7, the largest number of variables which can all have arbitrarily large powers and still multiply to an independent monomial is d. For any choice of d variables the remaining $n-d$ variables must all have bounded powers, and the total number of such monomials is $O(k^d)$. There are finitely many ways to choose d variables, so even if every such monomial were independent and each choice of d variables yielded a disjoint set of independent monomials, we would have at most $O(k^d)$. □

While this theorem tells us that the growth rate of $\dim P_{Q,F_k^n}(\mu)$ (as a function of k) is polynomial of degree d, it is not possible to give an upper bound on the leading coefficient of that polynomial. To see this, take the ideal generated by $x_{n-d}^t \cdot x_{n-d+1}^t \cdots x_n^t$ for some $t \in \mathbb{N}$. This ideal has Hilbert dimension d for any choice of t, but the number of monomials outside this ideal grows with t.

There remains the question of imposing additional moment vanishing conditions on a basis for $L^2_{Q,F_k^n,F_k^n}(\mu)$. We saw in example 7 that it is possible for an Alpert basis to satisfy extra orthogonality conditions beyond those directly imposed by the construction of the basis. However, in that example the basis Haar functions were piecewise constant and the accidental condition satisfied was orthogonality to x^2. In this section, we have restricted ourselves to considering monomials subject to a graded monomial order; this would disqualify the above example as the next monomial to be considered would be x, rather than x^2. We might hope that this extra restriction would cause accidental moment vanishing to not occur, and therefore allow the exact number of available orthognality conditions to be found by considering only the individual monomial spaces.

Sadly, in two or more dimenions this is still not the case. The following example shows that even with the additional restriction of a graded monomial order, it is possible for a Haar basis to contain accidental orthogonality.

Example 9. *Let μ be the measure with point masses located at the four points $(0.1, 0.1)$, $(0.4, 0.3)$, $(0.2, 0.7)$, and $(0.3, 0.8)$, each having mass $\frac{1}{4}$ and which has no mass elsewhere. Also let D^* be the standard dyadic grid, $Q \in D^*$ be the unit square, and Q_1 through Q_4 be the children of Q ordered counterclockwise from the upper right. Then the Haar space $L^2_{Q,\{1\},\{1\}}(\mu)$ has dimension one and $h_Q(x) = \mathbf{1}_{Q_2}(x) - \mathbf{1}_{Q_3}(x)$ is a Haar function for Q. This gives*

$$\begin{aligned}\int_Q h_Q(x) \cdot x \, d\mu(x) &= \int_Q \mathbf{1}_{Q_2}(x) \cdot x \, d\mu(x) - \int_Q \mathbf{1}_{Q_3}(x) \cdot x \, d\mu(x) \\ &= \frac{1}{4}\left(x\big|_{(0.2,0.7)} + x\big|_{(0.3,0.8)} - x\big|_{(0.1,0.1)} - x\big|_{(0.4,0.3)} \right) \\ &= \frac{1}{4}(0.5 - 0.5) \\ &= 0.\end{aligned}$$

So $h_Q(x)$ is orthogonal to x despite this not being imposed by the construction of the Haar basis.

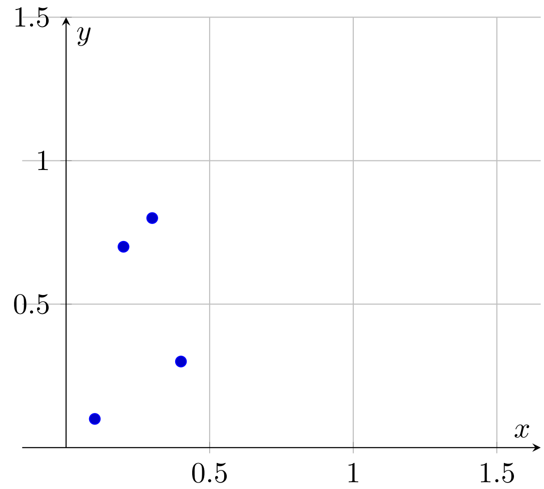

Figure 3.3.1: A point mass measure yielding accidental orthogonality in $\mathbb{R}^2$

The four point masses in this example were chosen to emphasize that μ need not contain any obvious geometric structure; as long as the x-coordinates of the point masses in each child sum to the same total $h_Q(x)$ will be orthogonal to x. This example also showcases that the extra orthogonality is possible even though 1 and x are linearly independent on each of Q, Q_2, and Q_3.

Interestingly, this behaviour is not replicated in one dimension. It turns out that in example 7, the choice of x^2 rather than x was crucial; a Haar basis in one dimension can never achieve orthogonality against x. We show this now.

Theorem 7. *Let μ be a locally finite positive Borel measure on $\mathbb{R}$, D be a dyadic grid, and $I \in D$ be an interval such that the Haar space $L^2_{I,\{1\},\{1\}}(\mu)$ is not trivial. Then the Haar function $h_I(x)$ on I is not orthogonal to x.*

Proof. A Haar function for I is given by:

$$h_I(x) = \frac{1}{\sqrt{\mu(I)}} \left(\sqrt{\frac{\mu(I_r)}{\mu(I_l)}} \mathbf{1}_{I_l}(x) - \sqrt{\frac{\mu(I_l)}{\mu(I_r)}} \mathbf{1}_{I_r}(x) \right).$$

Suppose toward a contradiction that $h_I(x)$ is orthogonal to x. Then we would have

$$\begin{aligned}\int_I h_I(x)\cdot x\,d\mu(x) &= \frac{1}{\sqrt{\mu(I)}}\int_I\left(\sqrt{\frac{\mu(I_r)}{\mu(I_l)}}\mathbf{1}_{I_l}(x)-\sqrt{\frac{\mu(I_l)}{\mu(I_r)}}\mathbf{1}_{I_r}(x)\right)\cdot x\,d\mu(x)\\ &= \frac{1}{\sqrt{\mu(I)}}\left(\sqrt{\frac{\mu(I_r)}{\mu(I_l)}}\int_{I_l}x\,d\mu(x)-\sqrt{\frac{\mu(I_l)}{\mu(I_r)}}\int_{I_r}x\,d\mu(x)\right)\\ &= 0\end{aligned}$$

and consequently

$$\frac{1}{\mu(I_l)}\int_{I_l}x\,d\mu(x)=\frac{1}{\mu(I_r)}\int_{I_r}x\,d\mu(x)\,.$$

Now suppose that we alter μ by scaling the measure of all measurable subsets of I_l by some positive constant k. Both $\mu(I_l)$ and $\int_{I_l}x\,d\mu(x)$ experience the same scaling factor, and so the lefthand side of the above equality remains unchanged. The same argument also applies to the righthand side, so without loss of generality we can make the simplifying assumption that $\mu(I_l)=\mu(I_r)=1$. We now have

$$\int_{I_l}x\,d\mu(x)=\int_{I_r}x\,d\mu(x)$$

where I_l and I_r are non-overlapping intervals of equal measure. This is impossible; since x is a monotone function the righthand integral must be larger than the lefthand. We have arrived at our contradiction, and we conclude that $h_I(x)$ cannot be orthogonal to x. □

3.3.2 Demonstrative Examples

At this point we have amassed a significant number of results regarding the properties of polynomial Alpert wavelets. In this section we demonstrate the application of these results to construct explicit bases for two example measures. The first is a point mass measure in one dimension, and the second is a twisted cubic in three dimensions. In each case we will use the standard dyadic grid.

Point Mass Measure in $\mathbb{R}$

Let $I = [0, 1)$ and suppose that μ assigns a mass of 1 to the points $x = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ and zero elsewhere. We will generate a basis for $L^2_{I,F_2^1,F_2^1}(\mu)$ with one basis element satisfying an additional quadratic orthogonality condition. Each of I_l and I_r contains two distinct point masses so P_{I_l,F_2^1}, P_{I_r,F_2^1} and P_{I,F_2^1} each have dimension 2—this is equivalent to saying that 1 and x are linearly independent on each of I_l, I_r, and I. Then by corollary 2 we have $\dim L^2_{I,F_2^1,F_2^1}(\mu) = 2+2-2 = 2$.

Take $\{\mathbf{1}_{I_l}, \mathbf{1}_{I_l}x, \mathbf{1}_{I_r}, \mathbf{1}_{I_r}x\}$ to be a starting basis for $L^2_{I,F_2^1,\varnothing}(\mu)$. Using Gram-Schmidt to orthonormalize this basis, we arrive at $B = \{b_1, b_2, b_3, b_4\}$ given by

$$\begin{aligned}
b_1 &= \frac{\sqrt{2}}{2}\mathbf{1}_{I_l} \\
b_2 &= \frac{\sqrt{2}}{2}\mathbf{1}_{I_l} - 4\sqrt{2}\cdot\mathbf{1}_{I_l}x \\
b_3 &= \frac{\sqrt{2}}{2}\mathbf{1}_{I_r} \\
b_4 &= \frac{5\sqrt{2}}{2}\mathbf{1}_{I_r} - 4\sqrt{2}\cdot\mathbf{1}_{I_r}x.
\end{aligned}$$

Notice that $\{b_1, b_2\}$ is an orthonormal basis for $P_{I_l,F_2^1}(\mu)$, and similarly that $\{b_3, b_4\}$ is an orthonormal basis for $P_{I_r,F_2^1}(\mu)$. Next we express the desired orthogonal functions in terms of B:

$$\begin{aligned}
\mathbf{1}_I &= \sqrt{2}\cdot b_1 + \sqrt{2}\cdot b_3 \\
\mathbf{1}_I x &= \frac{\sqrt{2}}{8}b_1 - \frac{\sqrt{2}}{8}b_2 + \frac{5\sqrt{2}}{8}b_3 - \frac{\sqrt{2}}{8}b_4 \\
\mathbf{1}_I x^2 &= \frac{\sqrt{2}}{32}b_1 - \frac{\sqrt{2}}{32}b_2 + \frac{13\sqrt{2}}{32}b_3 - \frac{5\sqrt{2}}{32}b_4.
\end{aligned}$$

Here we are taking advantage of the fact that $\mathbf{1}_I x^2 \in L^2_{I,F_2^1,\varnothing}(\mu)$. This will simplify future calculations, but we emphasize that this is a feature of this particular example and that additional orthogonality conditions cannot always be expressed this way.

Now to compute a basis for the orthogonal complement of $P_{I,F_2^1}(\mu)$ inside $L^2_{I,F_2^1,\varnothing}(\mu)$, we have:

$$P_{I,F_2^1}(\mu)^\perp = \text{Null}\begin{pmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ \frac{\sqrt{2}}{8} & -\frac{\sqrt{2}}{8} & \frac{5\sqrt{2}}{8} & -\frac{\sqrt{2}}{8} \end{pmatrix}_B \tag{3.3.1}$$
$$= \text{Span}\Big\{(1,-4,-1,0)^T,(0,1,0,-1)^T\Big\}_B.$$

Since $\mathbf{1}_I x^2 \in L^2_{I,F_2^1,\varnothing}(\mu)$ we can similarly apply the additional quadratic orthogonality condition by finding the orthogonal complement of $P_{I,F_3^1}(\mu)$ inside $L^2_{I,F_2^1,\varnothing}(\mu)$:

$$P_{I,F_3^1}(\mu)^\perp = \text{Null}\begin{pmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ \frac{\sqrt{2}}{8} & -\frac{\sqrt{2}}{8} & \frac{5\sqrt{2}}{8} & -\frac{\sqrt{2}}{8} \\ \frac{\sqrt{2}}{32} & -\frac{\sqrt{2}}{32} & \frac{13\sqrt{2}}{32} & -\frac{5\sqrt{2}}{32} \end{pmatrix}_B \tag{3.3.2}$$
$$= \text{Span}\left\{(1,-2,-1,-2)^T\right\}_B.$$

To construct a basis for $L^2_{I,F_2^1,F_2^1}(\mu)$ it suffices to select the unique basis element from 3.3.2 and any linearly independent basis element from 3.3.1—here either will suffice. For convenience we will select the basis given by $\left\{(1,-2,-1,-2)^T,(0,1,0,-1)^T\right\}_B$ as these elements happen to already be orthogonal. Normalizing, we arrive at an Alpert basis $\{a_1,a_2\}$ for $L^2_{I,F_2^1,F_2^1}(\mu)$ given by

$$a_1 = \left(\frac{\sqrt{5}}{10} - \frac{8\sqrt{5}}{5}x\right)\mathbf{1}_{I_l} + \left(\frac{11\sqrt{5}}{10} - \frac{8\sqrt{5}}{5}x\right)\mathbf{1}_{I_r}$$
$$a_2 = \left(\frac{1}{2} - 4x\right)\mathbf{1}_{I_l} + \left(-\frac{5}{2} + 4x\right)\mathbf{1}_{I_r}$$

where a_1 is additionally orthogonal to x^2.

Twisted Cubic in $\mathbb{R}^3$

Suppose that μ assigns arclength along the curve in $\mathbb{R}^3$ parameterized by

$$T = \left\{(t,t^2,t^3) : t \in \mathbb{R}\right\}$$

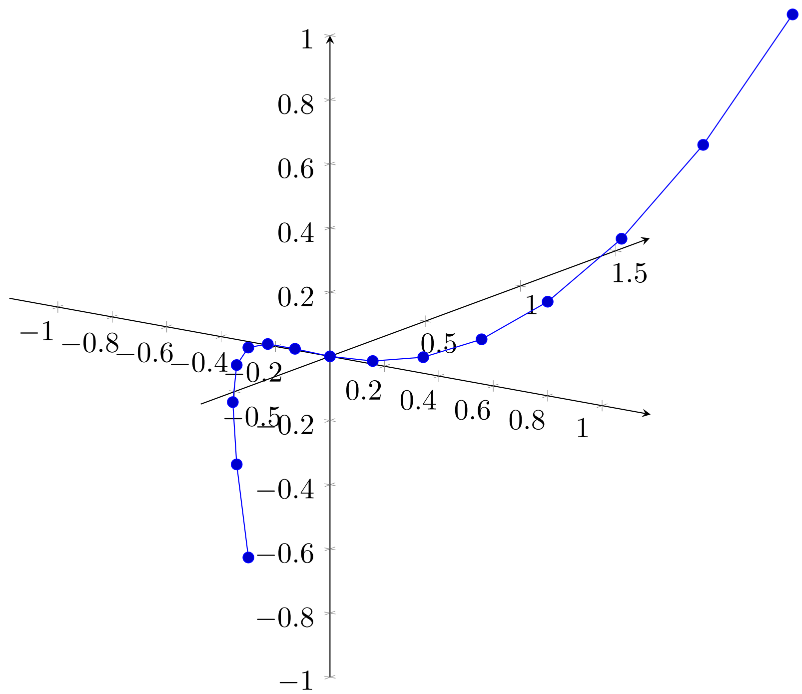

Figure 3.3.2: The twisted cubic in $\mathbb{R}^3$

and zero mass elsewhere. This curve is commonly called the twisted cubic. We see that T is the zero locus of $S = \{x^2 - y, x^3 - z\}$ so T is an algebraic set. Note that for any $Q \in D^*$, either T intersects Q in an arc or on a set of measure 0. Let $Q \in D^*$ such that $T \cap Q$ is an arc. Then we have $A_Q = T$ and $I_Q = \langle x^2 - y, x^3 - z\rangle$.

Since $-x(x^2-y)+(x^3-z) = xy-z \in I_Q$ but $xy \notin \langle x^2, x^3\rangle$ we see that S is not a Gröbner basis for I_Q. Similarly $y(x^2 - y) - x(xy - z) = xz - y^2 \in I_Q$ and $z(xy - z) - y(xz - y^2) = y^3 - z^2 \in I_Q$, so x^2, xy, xz, and y^3 must all be in the leading term ideal of I_Q. We also see $x^3 - z$ is extraneous as it can be written as a combination of $x^2 - y$ and $xy - z$, and $x^2 - y$ already contributes x^3 to the leading term ideal.

We now have $G = \{x^2 - y, xy - z, xz - y^2, y^3 - z^2\}$ as a candidate Gröbner basis. To confirm that is indeed a Gröbner basis, we compute the 6 S-polynomials in G (defined in Appendix B):

$$\begin{aligned} S(x^2 - y, xy - z) &= xz - y^2 \\ S(x^2 - y, xz - y^2) &= xy^2 - yz = y(xy - z) \\ S(x^2 - y, y^3 - z^2) &= x^2z^2 - y^4 = (xz + y^2)(xz - y^2) \\ S(xy - z, xz - y^2) &= y^3 - z^2 \\ S(xy - z, y^3 - z^2) &= xz^2 - y^2z = z(xz - y^2) \\ S(xz - y^2, y^3 - z^2) &= y^5 - xz^3 = y^2(y^3 - z^2) - z^2(xz - y^2) \end{aligned} \tag{3.3.3}$$

and we see that every S-polynomial reduces to 0 under multivariate division by G. Therefore G satisfies Buchberger's criterion and is a Gröbner basis for I_Q. Moreover, we will see that no monomial in G is divisible by another leading term in G, so G is the unique reduced Gröbner basis for I_Q (under degree lexicographic order). Then by theorem 5, a maximal set of independent monomials is given by precisely the monomials which are not multiples of any element of $LT(G)$:

Total degree	Independent monomials	Dependent monomials
0	1	-
1	x, y, z	-
2	y^2, yz, z^2	x^2, xy, xz
3	y^2z, yz^2, z^3	$x^3, x^2y, x^2z, xy^2, xyz, xz^2, y^3$
4	y^2z^2, yz^3, z^4	$x^4, x^3y, x^3z, x^2y^2, x^2yz, x^2z^2,$
		$xy^3, xy^2z, xyz^2, xz^3, y^4, y^3z$
$\vdots$	$\vdots$	$\vdots$

So for $k \in \mathbb{N}$ we have $\dim P_{Q,F_k^n}(\mu) = 3k - 2$, and the dimension of $L^2_{Q,F_k^n,\varnothing}(\mu)$ is $3k - 2$ times the number of children $Q'$of Q such that $Q' \cap T$ has positive measure.

This also verifies the growth rate given by theorem 6. From $G = \{x^2-y, xy-z, xz-y^2, y^3-z^2\}$ we see that $S = \{z\}$ satisfies the criteria for Hilbert dimension 1, and that no set of two variables does. So the ideal associated with the twisted cubic has Hilbert dimension 1—as we would expect from its geometry—and the number of independent monomials grows linearly with the total degree.

It is worth mentioning that this example may be slightly misleading in one respect: we were able to write down a linear function which gave $\dim P_{Q,F_k^n}(\mu)$ for all k, but this is not possible in general. We know from Theorem 6 that this dimension can be expressed as a polynomial for sufficiently large k, but the low order terms will not always follow this pattern. This didn't arise in the above example because the degrees in the Gröbner basis were small, and so the independent monomials settled into their eventual behaviour very quickly.

Chapter 4

Stability of Weighted Alpert Wavelets

In [19], Wilson shows that the Lebesgue Haar basis is stable in the following sense: if the elements of a Haar basis undergo small translations and dilations and an L^2 function f is then projected onto the resulting functions, the result is still close to f in the L^2 norm. In this section, we adapt this result to the setting of weighted Alpert wavelets on $\mathbb{R}$. We leave open the extension of these ideas to $\mathbb{R}^n$ due to difficulties which we will explain at the end of section 4.3.

4.1 Doubling Measures

Definition 17. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$. We say μ is doubling if there exists a constant $C_\mu > 0$ such that for any cube Q in $\mathbb{R}^n$ we have $\mu(2Q) \leq C_\mu \mu(Q)$, where $2Q$ is the cube with the same center as Q and double the side length.*

This is not the only definition for a doubling measure; it is also common to use balls instead of cubes. By considering the largest inscribed cube inside a ball and vice versa it can been seen that, other than a change in the constant C_μ, these definitions are equivalent. Lebesgue measure in $\mathbb{R}^n$ is, for instance, a doubling measure with doubling constant 2^n. Doubling measures can nevertheless exhibit unintuitively "bad" behaviour. For example:

- There exist doubling measures on $\mathbb{R}$ which are singular with respect to Lebesgue measure. [11]
- There exists a doubling measure on $\mathbb{R}^2$ which charges a rectifiable curve. [9]

In this chapter we are primarily interested in the convergence properties of wavelet bases when small perturbations are applied. While the above definition of a doubling measure is geometrically simple, it operates on a "large scale" which is poorly suited to this particular problem. Sawyer shows in [15, Lemma 4] that this definition can be translated into a "small scale" definition which looks at the annular halo around the boundary of a cube. Recall that for $k > 0$, $k \cdot Q$ is the cube with the same center as Q and with side length $k \cdot l(Q)$.

Lemma 8. *Suppose μ is a doubling measure on $\mathbb{R}^n$ and that Q is a cube in $\mathbb{R}^n$. Then for $0 < \delta < 1$ we have*

$$\mu\left((Q \setminus (1-\delta)Q\right) \leq \frac{C_\mu^2}{\log_2 \frac{1}{\delta}} \mu(Q), \;\; \textit{and}$$
$$\mu\left((1+\delta)Q \setminus Q\right) \leq \frac{C_\mu^2}{\log_2 \frac{1}{\delta}} \mu(Q).$$

Proof. Let $\delta = 2^{-m}$ and let $R \in C^{(m)}(Q)$ denote the set of m^{th}-level dyadic grandchildren, so each $R \in C^{(m)}(Q)$ has side length $l(R) = \delta \cdot l(Q)$. Define the collections

$$G^{(m)}(Q) = \left\{ R \in C^{(m)}(Q) : R \subset Q \text{ and } \partial R \cap \partial Q \neq \varnothing \right\}$$
$$H^{(m)}(Q) = \left\{ R \in C^{(m)}(Q) : 3R \subset Q \text{ and } \partial R \cap \partial(3Q) \neq \varnothing \right\}$$

Then

$$Q \setminus (1-\delta)Q = G^{(m)}(Q) \text{ and } (1-\delta)Q = \bigcup_{k=2}^{m} H^{(k)}(Q).$$

From the doubling condition we have $\mu(3R) \leq C_\mu^2 \mu(R)$ for all cubes R, so for $2 \leq k < m$ we have

$$\begin{aligned}
\mu\left(H^{(k)}(Q)\right) &= \sum_{R \in H^{(k)}(Q)} \mu(R) \\
&\geq \sum_{R \in H^{(k)}(Q)} \frac{1}{C_\mu^2} \mu(3R) \\
&= \frac{1}{C_\mu^2} \int_Q \left(\sum_{R \in H^{(k)}(Q)} \mathbf{1}_{3R} \right) d\mu \\
&\geq \frac{1}{C_\mu^2} \int_Q \left(\sum_{R \in H^{(k)}(Q)} \mathbf{1}_{R} \right) d\mu \\
&= \frac{1}{C_\mu^2} \mu\left(G^{(k)}(Q)\right) \\
&\geq \frac{1}{C_\mu^2} \mu\left(G^{(m)}(Q)\right) \\
&= \frac{1}{C_\mu^2} \mu(Q \setminus (1-\delta)Q).
\end{aligned}$$

Therefore

$$\mu(Q) \geq \sum_{k=2}^{m} \mu\left(H^{(k)}(Q)\right) \geq \frac{m-1}{C_\mu^2} \mu(Q \setminus (1-\delta)Q)$$

which completes the proof for the first inequality. The outer halo $(1+\delta)Q \setminus Q$ is just the inner halo for the cube $(1+\delta)Q$ (with a slightly different choice of δ) so the second inequality follows immediately. □

Conversely, if the annular halo around a cube is bounded in this way then μ is necessarily doubling.

Lemma 9. *Suppose μ is a locally finite positive Borel measure on $\mathbb{R}^n$, and suppose that $C > 0$ and $0 < \delta < 1$ are constants such that for any cube Q we have*

$$\mu\left((1+\delta)Q \setminus Q\right) \leq C\mu(Q)$$

Then μ is doubling.

Proof. We have $(1+\delta)Q = Q \cup ((1+\delta)Q \setminus Q)$, so by Lemma 8 we have

$$\mu\left((1+\delta)Q\right) \leq \mu(Q) + C\mu(Q) = (1+\delta)\,\mu(Q).$$

Since $(1+\delta)Q$ is also a cube, we can iteratively apply Lemma 8 to it. In particular we have $2Q \subseteq (1+\delta)^{\log_{(1+\delta)} 2} Q$, so after $\log_{(1+\delta)} 2$ applications of Lemma 8 we get

$$\mu(2Q) \leq (1+C)^{\log_{(1+\delta)} 2} \mu(Q).$$

Therefore μ is doubling with doubling constant at most $(1+C)^{\log_{(1+\delta)} 2}$. □

Informally, this halo condition ensures that a small translation or dilation of a cube can only produce a correspondingly small change in the cube's measure. The perturbations we consider in section 4.3 are defined using precisely these small translations.

4.2 Stability and Almost-Orthogonality

In [19, Section 2], Wilson shows stability (to be precisely defined momentarily) of one-dimensional Haar wavelets on $\mathbb{R}$ under small translations and dilations of the individual wavelets. Our goal in this section is to present that result and a generalized notion of perturbation in the more general context of arbitrary bases for $L^2(\mu)$, in preparation for proving the stability of Alpert bases in section 4.3.

Definition 18 (Perturbation)**.** *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$ and B be an orthonormal basis for $L^2(\mu)$. A perturbation P_η on B is a set of functions $\{p_b(\eta) : \mathbb{R}_+ \to L^2(\mu)\}_{b\in B}$ satisfying the conditions*

$$\lim_{\eta\to 0} \|b - p_b(\eta)\|_2 = 0$$

and

$$\lim_{\eta\to 0} |b(x) - p_b(\eta)(x)| = 0 \text{ for a.e. } x \in \mathbb{R}^n$$

for every $b \in B$.

Here η is the *perturbation parameter* of P_η. For a given value of $\eta > 0$, P_η gives a set of $L^2(\mu)$ functions $B^\eta = \{p_b(\eta)\}_{b\in B}$. We refer to this set as the perturbation of B under P_η, or just the perturbation of B by η when P_η is clear from context. We will write b^η to refer to the perturbation of a particular element $b \in B$.

Remark 5. *In this definition, the magnitude of a perturbation is given by a single parameter η. One might also be interested in perturbations which are intuitively defined by multiple parameters; for example, a perturbation could include both a translation and a dilation of the elements in B. Such cases can be handled by taking η to be the maximum among all parameters.*

Definition 19 (Stability)**.** *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$, B be an orthonormal basis for $L^2(\mu)$, and P be a perturbation on B. We say that B is stable under perturbation by P if there exists a positive function $\varphi(\eta) : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{\eta\to 0}\varphi(\eta) = 0$ such that for any $f \in L^2(\mu)$ and any $\eta > 0$ we have*

$$\|f - f^\eta\|_2 \leq \varphi(\eta)\|f\|_2,$$

where f^η is the projection of f onto B^η.

We emphasize that for stability, it is only the limiting behavior of the perturbation as η tends to zero that is important. For that reason we will generally assume that η is between zero and some fixed upper bound. Wilson showed that this notion of stability is closely related to *almost-orthogonality.*

Definition 20 (Almost-Orthogonal Set)**.** *A set $\{\psi_\gamma\}_{\gamma\in\Gamma} \subset L^2(\mu)$ is almost-orthogonal if there exists a constant $0 \leq C < \infty$ such that for all finite subsets $F \subset \Gamma$ and constants $\{\lambda_\gamma\}_{\gamma\in F} \subset \mathbb{R}$,*

$$\Big\| \sum_{\gamma\in F} \lambda_\gamma \psi_\gamma \Big\|_2 \leq C\Big(\sum_{\gamma\in F} |\lambda_\gamma|^2\Big)^{1/2}.$$

The almost-orthogonality constant for $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is the smallest constant C for which this inequality holds.

By duality, $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is almost-orthogonal with constant C if and only if for all $f \in L^2(\mu)$,

$$\Big(\sum_{\gamma\in\Gamma} |\langle f, \psi_\gamma\rangle|^2\Big)^{1/2} \leq C\|f\|_2$$

and C is the smallest constant for which this holds. This second formulation may be familiar to the reader by the name *Bessel sequence.*

Given a non-empty set Γ we can turn the collection of almost-orthogonal families indexed over Γ into a vector space $AO(\Gamma)$ by defining

$$\{\psi_\gamma\}_{\gamma\in\Gamma} + \{\phi_\gamma\}_{\gamma\in\Gamma} = \{\psi_\gamma + \phi_\gamma\}_{\gamma\in\Gamma}, \text{ and}$$
$$\alpha\{\psi_\gamma\}_{\gamma\in\Gamma} = \{\alpha\psi_\gamma\}_{\gamma\in\Gamma}.$$

Then the norm $\|\{\psi_\gamma\}_{\gamma\in\Gamma}\|_{AO(\Gamma)}$ is defined to be the almost-orthogonality constant of $\{\psi_\gamma\}_{\gamma\in\Gamma}$. An orthonormal basis for $L^2(\mu)$ is stable under perturbations which are $AO(\Gamma)$-small in the following sense:

Lemma 10. *Let $\delta > 0$ and $\{\psi_\gamma\}_{\gamma\in\Gamma} \subset L^2(\mu)$ be a complete orthonormal set. Suppose that $\{\widetilde{\psi}_\gamma\}_{\gamma\in\Gamma} \subset L^2(\mu)$ and $\{\widetilde{\psi}^*_\gamma\}_{\gamma\in\Gamma} \subset L^2(\mu)$ are two families such that $\big\|\{\psi_\gamma - \widetilde{\psi}_\gamma\}_{\gamma\in\Gamma}\big\|_{AO(\Gamma)}$ and $\big\|\{\psi_\gamma - \widetilde{\psi}^*_\gamma\}_{\gamma\in\Gamma}\big\|_{AO(\Gamma)}$ are both less than δ. For $f \in L^2(\mu)$, the series*

$$\sum_{\gamma\in\Gamma} \langle f, \widetilde{\psi}_\gamma\rangle \widetilde{\psi}^*_\gamma$$

converges to some $\widetilde{f} \in L^2(\mu)$, and $\|f - \widetilde{f}\|_2 \le \delta(2+\delta)\|f\|_2$.

Proof. First we observe that $\big\|\{\psi_\gamma\}_{\gamma\in\Gamma}\big\|_{AO(\Gamma)} = 1$ by the Pythagorean theorem, as $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is orthonormal. Then by the triangle inequality we have

$$\begin{aligned}
\big\|\{\widetilde{\psi}_\gamma\}_{\gamma\in\Gamma}\big\|_{AO(\Gamma)} &= \big\|\{\widetilde{\psi}_\gamma - \psi_\gamma + \psi_\gamma\}_{\gamma\in\Gamma}\big\|_{AO(\Gamma)} \\
&\le \big\|\{\psi_\gamma\}_{\gamma\in\Gamma}\big\|_{AO(\Gamma)} + \big\|\{\widetilde{\psi}_\gamma - \psi_\gamma\}\big\|_{AO(\Gamma)} \\
&\le 1 + \delta.
\end{aligned}$$

The same holds for $\{\widetilde{\psi}^*_\gamma\}_{\gamma\in\Gamma}$. Now we produce an error bound for $\widetilde{f}$.

$$\begin{aligned}
\|f - \widetilde{f}\|_2 &= \Big\| \sum_{\gamma\in\Gamma} \langle f, \psi_\gamma\rangle \psi_\gamma - \langle f, \widetilde{\psi}_\gamma\rangle \widetilde{\psi}^*_\gamma \Big\|_2 \\
&= \Big\| \sum_{\gamma\in\Gamma} \langle f, \psi_\gamma\rangle \psi_\gamma - \langle f, \psi_\gamma\rangle \widetilde{\psi}^*_\gamma + \langle f, \psi_\gamma\rangle \widetilde{\psi}^*_\gamma - \langle f, \widetilde{\psi}_\gamma\rangle \widetilde{\psi}^*_\gamma \Big\|_2 \\
&= \Big\| \sum_{\gamma\in\Gamma} \langle f, \psi_\gamma\rangle (\psi_\gamma - \widetilde{\psi}^*_\gamma) + (\langle f, \psi_\gamma\rangle - \langle f, \widetilde{\psi}_\gamma\rangle) \widetilde{\psi}^*_\gamma \Big\|_2 \\
&= \Big\| \sum_{\gamma\in\Gamma} \langle f, \psi_\gamma\rangle (\psi_\gamma - \widetilde{\psi}^*_\gamma) + \langle f, \psi_\gamma - \widetilde{\psi}_\gamma\rangle \widetilde{\psi}^*_\gamma \Big\|_2 \\
&\le \Big\| \sum_{\gamma\in\Gamma} \langle f, \psi_\gamma\rangle (\psi_\gamma - \widetilde{\psi}^*_\gamma) \Big\|_2 + \Big\| \sum_{\gamma\in\Gamma} \langle f, \psi_\gamma - \widetilde{\psi}_\gamma\rangle \widetilde{\psi}^*_\gamma \Big\|_2.
\end{aligned}$$

Here the last line is again by the triangle inequality. Now we can bound each of these norms using the almost-orthogonality constants derived above. For the first we have

$$\begin{aligned}\Big\|\sum_{\gamma\in\Gamma}\langle f,\psi_\gamma\rangle(\psi_\gamma-\widetilde{\psi}_\gamma^*)\Big\|_2 &\le \delta\Big(\sum_{\gamma\in\Gamma}\langle f,\psi_\gamma\rangle\Big)^{1/2}\\ &\le \delta\cdot 1\|f\|_2\\ &= \delta\|f\|_2.\end{aligned}$$

And for the second we have

$$\begin{aligned}\Big\|\sum_{\gamma\in\Gamma}\langle f,\psi_\gamma-\widetilde{\psi}_\gamma\rangle\widetilde{\psi}_\gamma^*\Big\|_2 &\le (1+\delta)\Big(\sum_{\gamma\in\Gamma}\langle f,\psi_\gamma-\widetilde{\psi}_\gamma\rangle\Big)^{1/2}\\ &\le (1+\delta)\cdot\delta\|f\|_2\\ &= (\delta+\delta^2)\|f\|_2.\end{aligned}$$

Combining these three calculations gives the desired result

$$\|f-\widetilde{f}\|_2 \le \delta\|f\|_2 + (\delta+\delta^2)\|f\|_2 = \delta(2+\delta)\|f\|_2.$$

□

Lastly, the almost-orthogonality condition reduces to a set of explicit integral calculations. This is the Schur test argument that Wilson gives in [19, Theorem 1].

Theorem 8. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$. Also let B be an orthonormal basis for $L^2(\mu)$ and P_η be a perturbation on B. Suppose that there exists a function $\varphi(\eta):\mathbb{R}_+\to\mathbb{R}_+$ with $\lim_{\eta\to 0}\varphi(\eta)=0$ such that for every $b_i\in B$ we have*

$$\sum_{b_j\in B}|\langle b_i-b_i^\eta,b_j\rangle|\le\varphi(\eta)$$

and for every $b_j\in B$

$$\sum_{b_i\in B}|\langle b_i-b_i^\eta,b_j\rangle|\le\varphi(\eta).$$

Then B is stable under pertubation by P_η.

Proof. Consider $\tau : \ell^2(B) \to \ell^2(B)$ defined initially for finite sums F by

$$\tau\left(\{\lambda_b\}_{b\in B}\right) := \left\{\sum_{b_i\in F} \lambda_b \left\langle b_i - b_i^\eta, b_j\right\rangle\right\}_{b_j\in B}.$$

By the Schur test, τ extends to a bounded operator with norm at most $\varphi(\eta)$ on all of $\ell^2(B)$. Since $\{b_j\}_{b_j\in B}$ is an orthonormal basis for $L^2(\mu)$ we get

$$\begin{aligned}\left\|\sum_{b_i\in F}\lambda_{b_i}(b_i - b_i^\eta)\right\|_2 &= \left(\sum_{b_j\in B}\left|\left\langle\sum_{b_i\in F}\lambda_{b_i}(b_i - b_i^\eta), b_j\right\rangle\right|^2\right)^{1/2}\\ &= \left(\sum_{b_j\in B}\left|\sum_{b_i\in F}\lambda_{b_i}\langle b_i - b_i^\eta, b_j\rangle\right|^2\right)^{1/2}\\ &= \|\tau(\{\lambda_{b_i}\}_{b_i\in F})\|_{\ell^2(B)}\\ &\le \varphi(\eta)\left(\sum_{b_i\in F}|\lambda_{b_i}|^2\right)^{1/2}.\end{aligned}$$

So we have $\left\|\sum_{b_i\in F}\lambda_{b_i}(b_i - b_i^\eta)\right\|_2 \le \varphi(\eta)\left(\sum_{b_i\in F}|\lambda_{b_i}|^2\right)^{1/2}$, and the family of functions $\{b_i - b_i^\eta\}_{b_i\in B}$ is almost-orthogonal in $L^2(\mu)$ with AO-constant at most $\varphi(\eta)$.By Lemma 10, B is stable under perturbation by P_η. □

This result reduces the question of stability to a set of relatively straightforward integral calculations. In particular these are easily checked for the Haar basis on $\mathbb{R}$ under small translations and dilations; each inner product $\langle b_i - b_i^\eta, b_j\rangle$ simplifies to the integral of a constant.

We note that the material in this section applies to any measure μ on $\mathbb{R}^n$ and any orthonormal basis for $L^2(\mu)$, though as we will see in section 4.3 this method has an important limitation. While these almost-orthogonality conditions are sufficient to guarantee stability of a basis, there do exist bases which are stable under certain perturbations but for which the sums in Theorem 8 do not converge.

4.3 Stability of Alpert Wavelets

The goal of this section is to prove that Alpert wavelet bases are stable under η-small translations when μ is a doubling measure. It will turn out that our technique is only sufficient for the task in one dimension, however most of the intermediate results we use hold in any dimension so we will give all but the final statement in full generality.

In this section we consider only Alpert spaces $L^2_{Q,F^n_k,F^n_k}(\mu)$, where each Alpert wavelet is composed piecewise of polynomials of degree less than k. Since there is no potential ambiguity, we will abbreviate this notation to $L^2_{Q,k,k}(\mu)$ for the sake of readability.

Definition 21. *A polynomial $p(x)$ in $\mathbb{R}^n$ is Q-normalized if*

$$\|\mathbf{1}_Q p\|_\infty = \sup_{x\in Q} |p(x)| = 1.$$

For a Q-normalized polynomial $p(x)$ we clearly have $\|\mathbf{1}_Q p\|_2 \leq \mu(Q)$. In [15] Sawyer shows that in a doubling measure μ a Q-normalized polynomial $p(x)$ cannot have an $L^2(\mu)$-norm that is very small relative to $\mu(Q)$.

Lemma 11. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$. If μ is doubling, then for every $k \in \mathbb{N}$ there exists a positive constant C_k such that*

$$\mu(Q) \leq C_k \int_Q |p(x)|^2 \, d\mu(x)$$

for all cubes Q in $\mathbb{R}^n$ and for all Q-normalized polynomials of degree less than k.

In fact the converse is also true: if the conclusion of Lemma 11 holds then μ is necessarily a doubling measure. The proof of this lemma is somewhat technical and not necessary to follow our work so we will omit it here, but the details can be found in [15, p. 16, lemma 20]. This result allows us to bound the ∞-norm of normalized functions in $L^2_{Q,k,0}(\mu)$.

Lemma 12. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$. If μ is doubling, then for any integer $k \geq 1$ there is a constant $C_k > 0$ such that for any cube Q in $\mathbb{R}^n$ and any $f \in L^2_{Q,k,0}(\mu)$ then*

$$\frac{\|f\|_\infty}{\|f\|_{L^2(\mu)}} \leq \sqrt{\frac{C_k}{\mu(Q)}}.$$

Proof. Let $Q' \in C(Q)$, $f \in L^2_{Q,k,0}(\mu)$, and consider $\mathbf{1}_{Q'}f$. This is a polynomial on Q', so by Lemma 11 we have

$$C_k \frac{\|\mathbf{1}_{Q'}f\|^2_{L^2(\mu)}}{\|\mathbf{1}_{Q'}f\|^2_\infty} = C_k \int_{Q'} \left| \frac{1}{\|\mathbf{1}_{Q'}f\|_\infty} \mathbf{1}_{Q'}f \right|^2 d\mu(x) \geq \mu(Q') > \mu(Q).$$

Rearranging gives

$$\frac{\|\mathbf{1}_{Q'}f\|_\infty}{\|\mathbf{1}_{Q'}f\|_{L^2(\mu)}} \leq \sqrt{\frac{C_k}{\mu(Q)}}$$

for each $Q' \in C(Q)$. This holds for every $Q' \in C(Q)$, and for at least one $Q' \in C(Q)$ we must have $\|\mathbf{1}_{Q'}f\|_\infty = \|f\|_\infty$. Lastly we have $\|\mathbf{1}_{Q'}f\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)}$, which gives

$$\frac{\|f\|_\infty}{\|f\|_{L^2(\mu)}} \leq \frac{\|\mathbf{1}_{Q'}f\|_\infty}{\|\mathbf{1}_{Q'}f\|_{L^2(\mu)}} \leq \sqrt{\frac{C_k}{\mu(Q)}}.$$

□

As an immediate consequence, every function in an Alpert basis for $L^2_{Q,k,k}(\mu)$ has ∞-norm at most $\sqrt{\frac{C_k}{\mu(Q)}}$. Recall that Haar wavelets are defined with a normalization factor $\frac{1}{\sqrt{\mu(Q)}}$, and since Haar functions are piecewise constant this factor directly determines the ∞-norm of the wavelet. The above result shows that this principle holds in general for higher-degree Alpert wavelets as well, with only the extra constant factor $\sqrt{C_k}$.

The next component we need is the continuity of the Gram-Schmidt algorithm. For two n-tuples of vectors U, U' define the distance between them as

$$\left\|U - U'\right\|_\mu := \sum_{i=1}^{n} \left\|\mathbf{u}_i - \mathbf{u}'_i\right\|_{L^2(\mu)}.$$

Lemma 13. *Suppose that V, V' are two n-tuples of linearly independent functions in $L^2(\mu)$, and that W, W' are the two outputs of applying the Gram-Schmidt algorithm to V, V'. Let $\widehat{V}$, $\widehat{V}'$, $\widehat{W}$, and $\widehat{W}'$ denote the corresponding tuples where each entry has been normalized with respect to the $L^2(\mu)$-norm. Then there is a continuous function $\nu : [0,\infty) \to [0,\infty)$ with $\nu(0) = 0$ such that*

$$\left\|\widehat{W} - \widehat{W}'\right\|_\mu \leq \nu\left(\left\|\widehat{V} - \widehat{V}'\right\|_\mu\right).$$

Proof. First we recall that the output of the Gram-Schmidt process does not depend on when the vectors are normalized, so without loss of generality assume that V, V' are sets of unit vectors. Given $m < n$ suppose that $\|\widehat{W} - \widehat{W}'\|_\mu \le \nu_m(\|\widehat{V} - \widehat{V}'\|_\mu)$ holds for any m-tuples V, V' and some continuous function $\nu_m : [0, \infty) \to [0, \infty)$ with $\nu_m(0) = 0$. Let $\mathbf{v}_{m+1}, \mathbf{v}'_{m+1} \in L^2(\mu)$ be functions with unit $L^2(\mu)$-norm which are not contained in $\operatorname{Span} V$, $\operatorname{Span} V'$ respectively.

Consider Gram-Schmidt applied to V appended with $\mathbf{v}_{m+1}$, and to V' appended with $\mathbf{v}'_{m+1}$. We get outputs

$$\mathbf{w}_{m+1} = \mathbf{v}_{m+1} - \sum_{i=1}^{m} \operatorname{proj}_{\mathbf{v}_i}(\mathbf{v}_{m+1})$$
$$\mathbf{w}'_{m+1} = \mathbf{v}'_{m+1} - \sum_{i=1}^{m} \operatorname{proj}_{\mathbf{v}'_i}(\mathbf{v}'_{m+1})$$

and we have

$$\begin{aligned}
&\left\|\widehat{W} \cup \{\widehat{\mathbf{w}}_{m+1}\} - \widehat{W}' \cup \{\widehat{\mathbf{w}}'_{m+1}\}\right\|_\mu \\
&= \left\|\widehat{W} - \widehat{W}'\right\|_\mu + \left\|\widehat{\mathbf{w}}_{m+1} - \widehat{\mathbf{w}}'_{m+1}\right\|_{L^2(\mu)} \\
&\le \nu_m\left(\left\|\widehat{V} - \widehat{V}'\right\|_\mu\right) + \left\|\widehat{\mathbf{w}}_{m+1} - \widehat{\mathbf{w}}'_{m+1}\right\|_{L^2(\mu)} \\
&\le \nu_m\left(\left\|\widehat{V} - \widehat{V}'\right\|_\mu\right) + \sum_{i=1}^{m+1} \left\|\widehat{\mathbf{v}}_i - \widehat{\mathbf{v}}'_i\right\|_{L^2(\mu)} \\
&= \nu_m\left(\left\|\widehat{V} - \widehat{V}'\right\|_\mu\right) + \left\|\widehat{V} \cup \{\widehat{\mathbf{v}}_{m+1}\} - \widehat{V}' \cup \{\widehat{\mathbf{v}}_{m+1}\}\right\|_\mu
\end{aligned}$$

where in the penultimate line we have used $\langle \widehat{\mathbf{v}}_i, \widehat{\mathbf{v}}_{m+1}\rangle \le 1$ for all $i = 1, ..., m$. The result then follows by induction on the lengths of V, V'. □

Lastly we need to show that the normalized monomials on Q and Q^η are close together, where $Q \in D$ is a dyadic cube in $\mathbb{R}^n$ and Q^η is a small translation of Q.

Theorem 9. *Let μ be a doubling measure on $\mathbb{R}^n$ with doubling constant C_μ. Let $Q \in D$ be a dyadic cube, $k \ge 1$ be an integer, and x^α be a monomial of degree less than k. Also let $0 \le \eta < \frac{1}{2}$, and $d \ge 2$ be an integer such that $2^{-d} < \eta \le 2^{1-d}$. Define Q^η to be the translation of Q by some vector with magnitude at most η. Finally let $g(x) := \widehat{x}^\alpha_Q - \widehat{x}^\alpha_{Q^\eta}$, where $\widehat{x}^\alpha_Q$ is the normalization of $\mathbf{1}_Q x^\alpha$. Then*

$$\|g\|_{L^2(\mu)} \le 3\sqrt{\frac{C_k C_\mu^2}{\ln \frac{1}{\eta}}}$$

where $C_k > 0$ is the constant determined in Lemma 12 which depends only on μ and k.

Proof. We will proceed by decomposing $Q \cup Q^\eta$ into $Q \cap Q^\eta$, $Q \setminus Q^\eta$, $Q^\eta \setminus Q$, and considering the restriction of g to each in turn. Beginning with $Q \cap Q^\eta$, we have

$$\begin{aligned}
\|\mathbf{1}_{Q\cap Q^\eta} g\|_{L^2(\mu)} &= \left\| \left(\frac{1}{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}} - \frac{1}{\|\mathbf{1}_{Q^\eta} x^\alpha\|_{L^2(\mu)}} \right) \mathbf{1}_{Q\cap Q^\eta} x^\alpha \right\|_{L^2(\mu)} \\
&= \left| \frac{\|\mathbf{1}_{Q\cap Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}} - \frac{\|\mathbf{1}_{Q\cap Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_{Q^\eta} x^\alpha\|_{L^2(\mu)}} \right| \\
&= \left| \frac{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)} - \|\mathbf{1}_{Q\setminus Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}} - \frac{\|\mathbf{1}_{Q^\eta} x^\alpha\|_{L^2(\mu)} - \|\mathbf{1}_{Q^\eta\setminus Q} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_{Q^\eta} x^\alpha\|_{L^2(\mu)}} \right| \\
&= \left| \frac{\|\mathbf{1}_{Q\setminus Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}} - \frac{\|\mathbf{1}_{Q^\eta\setminus Q} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_{Q^\eta} x^\alpha\|_{L^2(\mu)}} \right|. \\
&\leq \max\left(\frac{\|\mathbf{1}_{Q\setminus Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}}, \frac{\|\mathbf{1}_{Q^\eta\setminus Q} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_{Q^\eta} x^\alpha\|_{L^2(\mu)}} \right).
\end{aligned}$$

We claim that each of these ratios tends to 0 as $\eta \to 0$. By Lemma 12 we have

$$\frac{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_\infty} \geq \sqrt{\frac{\mu(Q)}{C_k}}.$$

The ratio of $L^2(\mu)$-norm to ∞-norm is maximized when x^α is constant, in which case the $L^2(\mu)$-norm is simply $\sqrt{\mu(Q)}$ times the ∞-norm. Combined with the above this gives

$$\sqrt{\frac{\mu(Q)}{C_k}} \leq \frac{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_\infty} \leq \sqrt{\mu(Q)}.$$

and therefore

$$\begin{aligned}
\frac{\|\mathbf{1}_{Q\setminus Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}} &\leq \frac{\frac{\|\mathbf{1}_{Q\setminus Q^\eta} x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_{Q\setminus Q^\eta} x^\alpha\|_\infty}}{\frac{\|\mathbf{1}_Q x^\alpha\|_{L^2(\mu)}}{\|\mathbf{1}_Q x^\alpha\|_\infty}} \\
&\leq \sqrt{\frac{C_k \mu(Q \setminus Q^\eta)}{\mu(Q)}}.
\end{aligned}$$

We now want to give an upper bound on the ratio $\frac{\mu(Q\setminus Q^\eta)}{\mu(Q)}$ in terms of C_μ and η. $Q \setminus Q^\eta$ is contained inside the interior halo of Q with width $\eta \cdot l(Q)$, so by Lemma 8 we have

$$\mu(Q \setminus Q^\eta) \leq \frac{C_\mu^2}{\ln \frac{1}{\eta}} \mu(Q).$$

This same argument applies with Q and Q^η reversed, with $Q \setminus Q^\eta$ contained inside the exterior halo of Q. We therefore conclude

$$\|\mathbf{1}_{Q \cap Q^\eta} g\|_{L^2(\mu)} \leq \sqrt{\frac{C_k C_\mu^2}{\ln \frac{1}{\eta}}}.$$

We now consider the restriction of g to $Q \setminus Q^\eta$. Here g is simply part of a normalized monomial on Q, so by Lemma 12 we have $\|\mathbf{1}_{Q \setminus Q^\eta} g\|_\infty \leq \sqrt{\frac{C_k}{\mu(Q)}}$. Then

$$\begin{aligned} \left\|\mathbf{1}_{Q \setminus Q^\eta} g\right\|_{L^2(\mu)} &\leq \left\|\mathbf{1}_{Q \setminus Q^\eta} g\right\|_\infty \cdot \sqrt{\mu(Q \setminus Q^\eta)} \\ &\leq \sqrt{\frac{C_k \mu(Q \setminus Q^\eta)}{\mu(Q)}} \end{aligned}$$

which is the same upper bound we computed previously for $\|\mathbf{1}_{Q \setminus Q^\eta} g\|_{L^2(\mu)}$. By symmetry we can use this same estimate for $Q^\eta \setminus Q$, and combining the three estimates we get

$$\|g\|_{L^2(\mu)} \leq 3\sqrt{\frac{C_k C_\mu^2}{\ln \frac{1}{\eta}}}.$$

□

Let us pause and take stock of our primary motivation: we want to apply Theorem 8 to show that an Alpert basis is stable under, for example, small translations. Because μ is in general not translation-invariant, translating the basis functions directly will result in functions which don't have the nice orthogonality properties of Alpert wavelets. One solution is to instead translate the underlying dyadic cubes, and then define the perturbed basis functions to simply be Alpert basis functions on the perturbed cube. This fixes the aforementioned problem, but introduces a new one: a given Alpert space has many possible bases, and we need specifically a basis on Q^η where each basis element is associated with and close to a basis element on Q. We can achieve this by selecting a canonical basis for $L^2_{Q,k,k}(\mu)$.

Let D be a dyadic grid on $\mathbb{R}^n$, $0 \leq \eta < \frac{1}{2}$, and $\{\eta_Q\}_{Q\in D}$ be a set of vectors in $\mathbb{R}^n$ with $0 \leq \|\eta_Q\|_\infty \leq \eta$. Define the set of perturbed cubes $\{Q^\eta\}_{Q\in D}$ to be the set of dyadic cubes each translated by $\eta_Q \cdot l(Q)$ where $l(Q)$ is the side length of Q. Recall that each $Q \in D$ has 2^n children, and that for fixed k there are $N_k := \binom{n+k-1}{k-1}$ monomials of degree less than k. Choose a canonical order on each of these two finite sets; let $Q_1, ..., Q_{2^n}$ denote the children of Q for arbitrary $Q \in D$ and let $x^{\alpha_1}, ..., x^{\alpha_{N_k}}$ denote the monomials of degree less than k, each according to the chosen order.

Take S to be the ordered tuple containing the restrictions of the monomials $x^{\alpha_1}, ..., x^{\alpha_{N_k}}$ first to Q, then to Q_1, then to Q_2, and so on until Q_{2^n}. S has the following key properties:

- S has $(2^n + 1)N_k$ entries, and each entry is the restriction of a monomial to a dyadic cube.
- S forms a spanning set (but not a basis) for the Alpert space $L^2_{Q,k,0}(\mu)$.
- The first N_k entries in S form a basis for the component space $P_{Q,k}(\mu)$.

Now apply the Gram-Schmidt algorithm to S and let $\widehat{S}$ denote the output tuple. The first N_k entries of $\widehat{S}$ still span $P_{Q,k}(\mu)$, and the final N_k entries are all zero as the monomials on any one child can be written as sums of monomials on Q and the monomials on each other child. The remaining $(2^n - 1)N_k := m$ middle entires are mutually orthogonal inside $L^2_{Q,k,0}(\mu)$ and are all orthogonal to the subspace $P_{Q,k}(\mu)$ inside $L^2_{Q,k,0}(\mu)$; in other words they form an orthonormal basis for $L^2_{Q,k,k}(\mu)$, the Alpert space of order k on Q. We let these entries in $\widehat{S}$ be the canonical choice of basis for $L^2_{Q,k,k}(\mu)$, and we denote them $a^Q_1, ..., a^Q_m$.

To a degree this is artificial; the Alpert space $L^2_{Q,k,k}(\mu)$ is the primary object of interest and clearly does not depend on the choice of basis. However this is also just a generalization of a detail that is often glossed over when discussing Haar wavelets. The Haar basis on $\mathbb{R}$ is not uniquely defined; the basis can be constructed with each wavelet either negative or positive on the left half. If we take a dyadic interval I and its Haar function h_I, then allow the perturbed h^η_I to be a Haar wavelet constructed with the opposite orientation, then $\lim_{\eta\to 0} h_I - h^\eta_I \neq 0$ and we see that our construction is not a valid perturbation.

Recall that we are trying to find estimates for all $|\langle g^Q_i, a^R_j\rangle|$ with $Q, R \in D$ and $1 \leq i, j \leq m$. We now have an estimate for the case where $Q = R$, so it remains to find estimates when $Q \neq R$.

By the orthogonality of Alpert bases, for $Q \neq R$ we have

$$\begin{aligned}\left|\langle g_i^Q, a_j^R\rangle\right| &= \left|\langle a_i^Q - h_i^{Q^\eta}, a_j^R\rangle\right| \\ &= \left|\langle a_i^Q, a_j^R\rangle - \langle a_i^{Q^\eta}, a_j^R\rangle\right| \\ &= \left|\langle a_i^{Q^\eta}, a_j^R\rangle\right|.\end{aligned}$$

We can therefore reverse the roles of Q and R by viewing R as the perturbation of some dyadic cube R^η inside a different dyadic grid D^η. If $l(Q) < l(R)$, then the resulting perturbation by $\eta_Q \cdot l(Q)$ is less than the maximum allowable $\eta \cdot l(R)$, and so is a valid perturbation.

Without loss of generality suppose that $l(Q) \geq l(R)$. Let $m \geq 0$ be the integer such that $2^{-m} l(Q) < l(R) \leq 2^{1-m} l(Q)$. There are two cases to consider: either R and Q are disjoint, or R is contained in a child of Q. In the former case, by Lemma 6 we have

$$\begin{aligned}\left|\langle a_i^{Q^\eta}, a_j^R\rangle\right| &\leq \|a_i^{Q^\eta}\|_\infty \|h_j^R\|_\infty \cdot \mu(Q^\eta \cap R) \\ &\leq \sqrt{\frac{C_k}{\mu(Q^\eta)}}\sqrt{\frac{C_k}{\mu(R)}} \cdot \mu(Q^\eta \cap R) \\ &\leq \frac{C_k \mu(Q^\eta \cap R)}{\sqrt{\mu(Q^\eta)\mu(R)}}.\end{aligned}$$

Now instead suppose that R is contained inside Q. As an Alpert wavelet, a_j^R is orthogonal to polymials of degree less than k, so our inner product is only nonzero if R overlaps the boundary of some child Q' of Q^η. The restriction $\mathbf{1}_{Q'} a_i^{Q^\eta}$ is a polynomial of degree less than k, so if we were to extend its domain to include all of $Q' \cup R$ then it would be orthogonal to a_j^R. Consequently

$$\begin{aligned}\left|\langle a_i^{Q^\eta}, a_j^R\rangle\right| &\leq \|a_i^{Q^\eta}\|_\infty \|a_j^R\|_\infty \cdot \mu(R \setminus Q') \\ &\leq \sqrt{\frac{C_k}{\mu(Q^\eta)}}\sqrt{\frac{C_k}{\mu(R)}} \cdot \mu(R \setminus Q') \\ &\leq \frac{C_k \mu(R \setminus Q')}{\sqrt{\mu(Q^\eta)\mu(R)}}.\end{aligned}$$

We need to bound these ratios in two different ways. We have $\mu(R \setminus Q') \leq \mu(R)$, which gives a good bound when R is small relative to Q. We also have $\mu(R \setminus Q') \leq \frac{C_\mu^2}{\ln \frac{1}{\eta}} \mu(Q^\eta)$ by Lemma

8, which gives a good bound when R and Q are similar in size. These of course also apply to $\mu(Q^\eta \setminus R)$. Using the first inequality we get

$$\left|\langle a_i^{Q^\eta}, a_j^R\rangle\right| \leq C_k\sqrt{\frac{\mu(R)}{\mu(Q^\eta)}}$$

and using the second we get

$$\left|\langle a_i^{Q^\eta}, a_j^R\rangle\right| \leq \frac{C_kC_\mu^2}{\ln\frac{1}{\eta}}\sqrt{\frac{\mu(Q^\eta)}{\mu(R)}}.$$

Lastly we note that Q^η always contains at least one of the children of Q and is always contained in $2Q$, so we restate these inequalities in terms of $\mu(Q)$:

$$\left|\langle a_i^{Q^\eta}, a_j^R\rangle\right| \leq \frac{C_kC_\mu}{\sqrt{C_\mu^2 - 2^n + 1}}\sqrt{\frac{\mu(R)}{\mu(Q)}},$$

$$\left|\langle a_i^{Q^\eta}, a_j^R\rangle\right| \leq \frac{C_kC_\mu^{5/2}}{\ln\frac{1}{\eta}}\sqrt{\frac{\mu(Q)}{\mu(R)}}.$$

To use Theorem 8, we need to show that for every $R \in D$ and every $1 \leq j \leq m$

$$\sum_{Q\in D}\sum_{1\leq i\leq m}\left|\langle g_i^Q, a_j^R\rangle\right| \leq \varphi(\eta)$$

and for every $Q \in D$ and $1 \leq i \leq m$

$$\sum_{R\in D}\sum_{1\leq j\leq m}\left|\langle g_i^Q, a_j^R\rangle\right| \leq \varphi(\eta)$$

where $m := (2^n - 1)N_k$ is the dimension of $L^2_{Q,k,k}(\mu)$. However as these inner products reduce to $\langle a_i^{Q^\eta}, a_j^R\rangle$ whenever $Q \neq R$, every sum of the second kind can be interpreted as a sum of the first kind with a different underlying dyadic grid. It will therefore suffice to produce an estimate for the first sum. For $Q = R$, Theorem 9 gives

$$\|\hat{x}_Q^\alpha - \hat{x}_{Q^\eta}^\alpha\|_{L^2(\mu)} \leq 3\sqrt{\frac{C_kC_\mu^2}{\ln\frac{1}{\eta}}}.$$

The canonical basis for $L^2_{Q,k,k}(\mu)$ is the output of Gram-Schmidt applied to N_k monomials on each of Q and all but one of Q's 2^n children. So by the Cauchy-Schwarz inequality and Theorem 9, we have

$$\begin{aligned}\sum_{1\leq i\leq m}\left|\langle g_i^Q, a_j^Q\rangle\right| &\leq \sum_{1\leq i\leq m}\left\|g_i^Q\right\|_{L^2(\mu)}\cdot\left\|a_j^Q\right\|_{L^2(\mu)}\\ &\leq m\cdot\nu\left(3\sqrt{\frac{C_kC_\mu^2}{\ln\frac{1}{\eta}}}\right).\end{aligned}$$

Now suppose that $Q\neq R$ and $2^t l(Q)=l(R)$ for some integer $t\geq 0$. The only cubes $Q\neq R$ which can result in a non-zero inner product are those which share at least one boundary point with a child of R; let Ω denote the set of all such Q. We want to split the following sum at some index $T(\eta)$ to use each of our two estimates. This gives

$$\begin{aligned}&\sum_{t\geq 0}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}}\sum_{1\leq i\leq m}\left|\langle g_i^Q, h_j^R\rangle\right|\\ &\leq \sum_{0\leq t<T(\eta)}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}} m\left|\langle g_i^Q, h_j^R\rangle\right| + \sum_{t\geq T(\eta)}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}} m\left|\langle g_i^Q, h_j^R\rangle\right|\\ &\leq \sum_{0\leq t<T(\eta)}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}}\frac{mC_kC_\mu^{5/2}}{\ln\frac{1}{\eta}}\sqrt{\frac{\mu(R)}{\mu(Q)}} + \sum_{t\geq T(\eta)}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}}\frac{mC_kC_\mu}{\sqrt{C_\mu^2-2^n+1}}\sqrt{\frac{\mu(Q)}{\mu(R)}}\\ &= \frac{mC_kC_\mu^{5/2}}{\ln\frac{1}{\eta}}\sum_{0\leq t<T(\eta)}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}}\sqrt{\frac{\mu(R)}{\mu(Q)}} + \frac{mC_kC_\mu}{\sqrt{C_\mu^2-2^n+1}}\sum_{t\geq T(\eta)}\sum_{\substack{Q\in\Omega\\ 2^t l(Q)=l(R)}}\sqrt{\frac{\mu(Q)}{\mu(R)}}.\end{aligned}$$

We will refer to the two double sums in this expression as (I) and (II) respectively. Note that for a fixed choice of $t\geq 0$, the union of all $Q\in\Omega$ is presicely the union of the interior and exterior halos of width $2^{-t}l(R)$ of each child of R. To see why this splitting was necessary, recall that we are trying to construct an upper bound in terms of η that will vanish as $\eta\to 0$. The estimate in (I) alone is insufficient; because R is fixed and Q varies over all cubes smaller than R, even the individual terms $\sqrt{\frac{\mu(R)}{\mu(Q)}}$ will diverge. On the other hand, while the estimate in (II) does avoid this problem it does not actually depend on η and therefore does not give the vanishing that we need. The solution is to construct $T(\eta)$ so that terms from (II) are slowly absorbed into (I) as η decreases.

As was foretold, this is the point at which the proof fails in higher dimensions. While the individual terms $\sqrt{\frac{\mu(Q)}{\mu(R)}}$ do approach zero as the size of Q decreases, the number of cubes Q in the halo around R also grows in any dimension higher than 1. The full details require some setup and we are close to the end of our current task, so we will finish proving stability of Alpert wavelets in one dimension and then present the failure in higher dimenions in section 4.3.1.

In addition to what came before, now assume that μ is a doubling measure on $\mathbb{R}$. Instead of referring to cubes $Q, R \in D$ we will use intervals $I, J \in D$ to emphasize the distinction. Next we must give the construction of $T : (0, \frac{1}{2}) \to \mathbb{N}$, the breakpoint between the two above double sums. Note that for fixed values of $T(\eta)$, (I) is finite and so tends to 0 as $\eta \to 0$. For $a \in \mathbb{N}$ let $f(a, \eta)$ denote the value of (I) that arises from fixing $T(\eta) = a$. Each $f(a, \eta)$ is therefore a function that tends to 0 as $\eta \to 0$. Define $T(\eta)$ to be the largest index a such that $f(a, \eta) < 2^{-a}$, to a minimum of 1 if no index satisfies this condition. Since each $f(a, \eta)$ decreases to 0 we have $T(\eta) \to \infty$ as $\eta \to 0$, and so by construction we have (I) $\to 0$ as $\eta \to 0$. Then since each $I \in \Omega$ can appear in only one of (I) and (II), and every I appears in (I) for $T(\eta)$ sufficiently large, it follows that (II) also tends to 0 as $\eta \to 0$.

Now instead suppose that $2^t l(I) = l(J)$ for some integer $t < 0$. There are at most 2 intervals I of length $2^{-t} l(J)$ which will yield non-zero inner products; let Ω be the set containing all such I. As before we sum over all $t < 0$ and split at a value $T(\eta)$ so that both parts of the sum vanish as $\eta \to 0$.

$$\begin{aligned}
&\sum_{t<0} \sum_{\substack{I \in \Omega \\ 2^t l(I) = l(J)}} \sum_{1 \le i \le m} \left| \langle g_i^I, a_j^J \rangle \right| \\
\le &\sum_{T(\eta) \le t < 0} \sum_{\substack{I \in \Omega \\ 2^t l(I) = l(J)}} m \left| \langle g_i^I, a_j^J \rangle \right| + \sum_{t < T(\eta)} \sum_{\substack{Q \in \Omega \\ 2^t l(I) = l(J)}} m \left| \langle g_i^I, a_j^J \rangle \right| \\
\le &\sum_{T(\eta) \le t < 0} \sum_{\substack{I \in \Omega \\ 2^t l(I) = l(J)}} \frac{m C_k C_\mu^{5/2}}{\ln \frac{1}{\eta}} \sqrt{\frac{\mu(I)}{\mu(J)}} + \sum_{t < T(\eta)} \sum_{\substack{I \in \Omega \\ 2^t l(I) = l(J)}} \frac{m C_k C_\mu}{\sqrt{C_\mu^2 - 2^n + 1}} \sqrt{\frac{\mu(J)}{\mu(I)}}
\end{aligned}$$

As Ω contains no more intervals than in the previous case, we can reuse the construction of $T(\eta)$ to force both sums to vanish as $\eta \to 0$. We therefore have a monotone function $\varphi(\eta) : [0, \frac{1}{2}) \to \mathbb{R}_+$

with $\lim_{\eta\to 0}\varphi(\eta) = 0$ such that for every $J \in D$ and every $1 \le j \le m$

$$\sum_{J\in D}\sum_{1\le i\le m} \left|\langle g_i^I, a_j^J\rangle\right| \le \varphi(\eta)$$

and for every $Q \in D$ and $1 \le i \le m$

$$\sum_{I\in D}\sum_{1\le j\le m} \left|\langle g_i^I, a_j^J\rangle\right| \le \varphi(\eta)$$

so by direct application of Theorem 8 we arrive at the desired stability theorem.

Theorem 10. *Let μ be a doubling measure on $\mathbb{R}$, D be a dyadic grid, and B be the canonical Alpert basis constructed over D. Also let P be a perturbation applied to B defined as follows: for each interval $I \in D$ and each wavelet $b \in B$ associated with I, the perturbation of b by η is defined to be the equivalent canonical Alpert wavelet on I^η, where I^η is a translation of I by at most $\eta \cdot l(I)$. Then B is stable under the perturbation P.*

4.3.1 Stability in Higher Dimensions

A natural thought is to try to extend the above result to any Alpert basis on an n-dimensional doubling measure, since all but the final computation were presented in an arbitrary dimension. Alas, as we alluded to earlier this is not possible. This observation appears in [19, appendix 1], but it is sufficiently important that we present our own explanation here.

The rather dense notation in the preceding result obscures the problem, so let us instead consider the much simpler case of a 2-dimensional Haar basis in Lebegue measure. As before, let D be a dyadic grid on $\mathbb{R}^2$ and let B be a Haar basis for $L^2(\mathbb{R}^2)$ defined on D. Choose a cube $Q \in D$ and define $\widetilde{D}$ to be the subset of all cubes which are smaller than Q, disjoint from Q, and which have a boundary face contained in ∂Q. Let P_η be the perturbation which takes every cube $R \in \widetilde{D}$ and translates it along an axis direction by $\eta \cdot l(R)$ in the direction of Q. Assuming that $\eta < \frac{1}{2}$, each of the perturbed cubes R^η now overlaps Q and the area of the overlap is precisely $\eta \cdot \mu(R)$.

Now for every dyadic cube there are three associated Haar functions in $\mathbb{R}^2$; choose one of the Haar functions on Q to be h_Q. Similarly for each $R \in \widetilde{D}$ there are three Haar functions on R^η; we claim at least one of them must be not orthogonal to h_Q. To see this, note that the restriction of h_Q to R^η is constant, and that two of the children of R^η are still disjoint from Q. We know that $L^2_{R^\eta,1,1}(\mathbb{R}^2)$ contains, for example, a Haar function which is positive on the two children which overlap Q and negative on the two that don't. This Haar function is not orthogonal to h_Q, so even if it is not one of the Haar functions selected for the basis B there must be at least one Haar function in B which is not orthogonal to h_Q. Let this basis function be h^η_R; since we are working in Lebesgue measure, h^η_R is just the translation of the equivalent Haar function h_R by $\eta \cdot l(R)$ in the direction of Q.

To use the Schur test argument from Theorem 8, it is necessary (though not sufficient) that the following sum converges to a finite value controlled by η:

$$\sum_{R\in\widetilde{D}} \left|\langle h^\eta_R, h_Q\rangle\right|.$$

Since we are working in Lebesgue measure, each h^η_R has ∞-norm $\frac{1}{\sqrt{\mu(R^\eta)}} = \frac{1}{\sqrt{\mu(R)}}$. Similarly $\|h_Q\|_\infty = \frac{1}{\sqrt{\mu(Q)}}$. Both h_Q and h^η_R are constant on $Q\cap R^\eta$ and we know the area of $Q\cap R^\eta$, so we have

$$\left\langle h^\eta_R, h_Q\right\rangle = \frac{1}{\sqrt{\mu(R)}}\cdot\frac{1}{\sqrt{\mu(Q)}}\cdot\eta\mu(R) = \eta\cdot\sqrt{\frac{\mu(R)}{\mu(Q)}}.$$

Therefore

$$\sum_{R\in\widetilde{D}} \left|\langle h^\eta_R, h_Q\rangle\right| = \eta\cdot\sum_{R\in\widetilde{D}}\sqrt{\frac{\mu(R^\eta)}{\mu(Q)}}.$$

Now $k\geq 1$ be an integer and consider all the cubes $R\in\widetilde{D}$ which have length $\frac{1}{2^k}$. There are $4{\cdot}2^k$ such cubes; 2^k along each of the four sides of Q. Also for such cubes we have $\mu(R^\eta) = \frac{1}{2^{2k}}\mu(Q)$.

We can use this to decompose our sum by cube size:

$$\begin{aligned}\sum_{R\in\widetilde{D}} \left|\langle h_R^\eta, h_Q\rangle\right| &= \eta \cdot \sum_{R\in\widetilde{D}} \sqrt{\frac{\mu(R)}{\mu(Q)}} \\ &= \sum_{k=1}^{\infty} 4\cdot 2^k \cdot \eta \sqrt{\frac{\frac{1}{2^{2k}}\mu(Q)}{\mu(Q)}} \\ &= 4\eta \cdot \sum_{k=1}^{\infty} 2^k \sqrt{\frac{1}{2^{2k}}} \\ &= 4\eta \sum_{k=1}^{\infty} 1.\end{aligned}$$

So even just considering one of the three basis elements for each dyadic cube, the sums needed for Theorem 8 already diverge. This does not mean that the Haar basis is unstable under translations in $\mathbb{R}^2$—Wilson proved otherwise in [19]—but it means that the technique we used to prove 8 is insufficient to handle the higher-dimensional cases directly.

4.4 Instability in Non-Doubling Measures

Our result from section 4.3 leads to a natural question: if the doubling property is sufficient to ensure that Alpert bases are stable, then is doubling also necessary for stability? Or could it be that some non-doubling measures also enjoy a similar stability property?

To approach this question, we first need to emphasize a point that has so far been largely ignored: in non-doubling measures, whether an Alpert basis is stable under a given perturbation may depend on the underlying choice of dyadic grid.

Example 10. *Let μ be the measure on $\mathbb{R}$ which assigns point masses of weight 1 at points $x = \frac{1}{4}$ and $x = \frac{3}{4}$, and zero mass elsewhere. Let D^* be the standard dyadic grid. Then $L^2(\mu)$ is a 2-dimensional vector space, and a Haar basis over D^* is given by the constant function 1 and the Haar wavelet $h^{[0,1)}$.*

Let I be a translation of $[0,1)$ by some $\eta \in \mathbb{R}$. Clearly the constant function is unchanged under this perturbation. The perturbed Haar wavelet h^I is exactly equal to $h^{[0,1)}$ if $\frac{1}{4} \in I_l$ and $\frac{3}{4} \in I_r$, and is identically zero otherwise. These two inclusions are satisfied if and only if $|\eta| < \frac{1}{4}$.

Since the Haar basis over D^* is entirely unchanged for perturbations $|\eta| < \frac{1}{4}$, it satisfies our definition of stability.

Now take D to be a translation of D^* by $\frac{1}{4}$. A Haar basis over D is given by the constant function 1 and the Haar wavelet $h^{[\frac{1}{4},\frac{5}{4})}$. Take I to be a translation of $[\frac{1}{4},\frac{5}{4})$ by any $\eta > 0$. Since we have $\mu(I_r) = 0$, the resulting Haar wavelet for I is identically zero. As this holds for any $\eta > 0$, the Haar basis over D does not satisfy our definition of stability.

So even highly non-doubling measures can be stable for the right choice of dyadic grid. This shows that precise formulation of the question should be: for a given measure μ, are Alpert bases stable for *every* dyadic grid?

Conjecture 1. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^n$. Suppose that μ has the property that, for any dyadic grid D and for any Alpert basis A defined on D, A is stable under small translations in the same sense as in theorem 10. Then μ must be a doubling measure.*

This result, if true, would give a characterization of doubling measures as precisely those in which Alpert bases are stable under small translations. To illustrate the difficulty in approaching the problem, suppose we take a non-doubling measure μ on $\mathbb{R}$. To be non-doubling means that, for any constants $C, \eta > 0$, there is some region somewhere in $\mathbb{R}$ where translating some interval I by a factor of η causes the measure of I to change by at least a factor of C. To use our approach from section 4.3, we would try to show that a corresponding Alpert wavelet on I must experience a similarly large change in L^2-norm.

The difficulty is that, for small η, the non-doubling assumption only provides information about the exterior regions of I, namely those regions which either enter or leave I during the translation. Alpert wavelets, in contrast, depend heavily on the structure of μ in the interior of I.

Despite this difficulty, we present a partial result which leans in the direction of 1 being true. In the following theorem we assume that the measure charges an infinite amount of mass toward both positive infinity and negative infinity, so that our bases do not need to include functions on the dyadic tops. This eliminates some trivial cases like a single point mass measure, which is technically stable for all dyadic grids under our definitions but only for the uninteresting reason that all Haar functions in such a measure are the zero function.

Theorem 11. *Let μ be a locally finite positive Borel measure on $\mathbb{R}$ such that $\int_{-\infty}^{0} d\mu\, x = \int_{0}^{\infty} d\mu\, x = \infty$ and suppose that μ contains a point mass at $x_0 \in \mathbb{R}$. Then there exist dyadic grids D for which a Haar basis B defined on D is unstable under translations.*

Proof. Let I be some finite interval which has x_0 as its right endpoint, and such that the interior of I_l has positive measure; by the assumption on μ such an I is guaranteed to exist. Let I^η denote the translation of I to the right by $\eta \cdot l(I)$ for a perturbation parameter $0 < \eta < \frac{1}{2}$; note that $x_0 \in I^\eta$. Let B be the Haar basis for some dyadic grid containing I, and let B^η be the perturbed Haar basis containing all the same functions as B, except the Haar function on I has been replaced by the Haar function on I^η. Lastly, let f be the indicator function on I.

Recall that, if B is stable under this perturbation, then we must have

$$\|f - f^\eta\|_2 \leq \varphi(\eta)\|f\|_2$$

where f^η denotes projection onto B^η and $\varphi(\eta)$ is some positive function that vanishes as $\eta \to 0$. From this construction, we have the following:

1. $\|f\|_2 = \sqrt{\mu(I)}$

2. $\langle f, h^I \rangle = 0$ by the moment vanishing of h^I.

Since B and B^η differ only at I, we can compute

$$\begin{aligned} \|f - f^\eta\|_2 &= \|\langle f, h^I \rangle h^I - \langle f, h^{I^\eta} \rangle h^{I^\eta}\|_2 \\ &= \|0 - \langle f, h^{I^\eta} \rangle h^{I^\eta}\|_2 \\ &= |\langle f, h^{I^\eta} \rangle| \|h^{I^\eta}\|_2 \\ &= |\langle f, h^{I^\eta} \rangle|. \end{aligned}$$

Now if we were to extend f to be an indicator on all of $I \cup I^\eta$ then f and h^{I^η} would be orthogonal by the moment vanishing of h^{I^η}. That means

$$\int_I h^{I^\eta}\, d\mu(x) + \int_{I^\eta \setminus I} h^{I^\eta}\, d\mu(x) = 0$$

and so

$$|\langle f, h^{I^\eta}\rangle| = \left|\int_{I^\eta \setminus I} h^{I^\eta}\, d\mu(x)\right|.$$

Since h^{I^η} is constant on $I^\eta \setminus I$, we can simplify this futher. Recall that h^{I^η} has formula

$$h^{I^\eta}(x) = \frac{1}{\sqrt{\mu(I^\eta)}}\left(\sqrt{\frac{\mu(I^\eta_r)}{\mu(I^\eta_l)}}\mathbf{1}_{I^\eta_l}(x) - \sqrt{\frac{\mu(I^\eta_l)}{\mu(I^\eta_r)}}\mathbf{1}_{I^\eta_r}(x)\right)$$

which gives

$$\left|\int_{I^\eta \setminus I} h^{I^\eta}\, d\mu(x)\right| = \mu(I^\eta \setminus I)\cdot\sqrt{\frac{\mu(I^\eta_l)}{\mu(I^\eta)\mu(I^\eta_r)}}.$$

We now need to argue that this quantity cannot vanish as $\eta \to 0$. Clearly $\mu(I^\eta \setminus I)$ is non-vanishing since $\mu(I^\eta \setminus I) \geq \mu(x) > 0$. The denominator $\mu(I^\eta)\mu(I^\eta_r)$ cannot grow arbitrarily large since μ is locally finite. All that remains is $\mu(I^\eta_l)$, which clearly cannot vanish as $\eta \to 0$ if there is any mass in the interior of I_l. We initially chose I specifically to satisfy this criterion, and therefore $\|f - f^\eta\|_2$ cannot vanish as $\eta \to 0$. We conclude that any dyadic grid containing I will have a Haar basis that is unstable under translations. $\square$

Chapter 5

Conclusion

In this thesis we presented two main contributions to the field of wavelet analysis. The first is our work in chapter 3, which gives a thorough description of how the dimensions of Alpert spaces are determined by the geometry of the measure. In the particular case of Alpert bases constructed from polynomials we found that, via a Gröbner basis technique, information about the geometry of the measure is sufficient to determine the size of Alpert bases of any degree.

Our second primary contribution is the work in section 4.3, culminating in theorem 10. Specifically we showed that Alpert bases are stable under small translations of the underlying dyadic intervals in a doubling measure on $\mathbb{R}$. This result was achieved by adapting techniques developed by Wilson [19] to the measure-theoretic setting.

This work also presents a concise introduction to the topic of weighted Haar and Alpert wavelets. The material in chapter 2 can be read with only a standard analysis background and some introductory measure theory, which we hope renders the subject accessible to readers who do not themselves have a background in wavelet analysis.

5.1 Further Questions

In chapter 3 we described the geometric structure of Alpert wavelet bases over arbitrary measures, and this project is now largely complete. One potential avenue for improvement is in the statement of theorem 3, in which each additional moment vanishing condition applied to a basis reduces the overall dimension of the Alpert space either by 0 or by 1. We were not able to identify a technique for determining in advance (i.e. using only the structure of the measure) which result a particular moment vanishing condition will yield, and we gave an example demonstrating that “similar looking” measures can yield different results. Nevertheless, it remains possible that such a technique does exist and merely eluded us.

Likely of more interest are the moment vanishing conditions themselves. As was mentioned in section 2.4, Alpert bases are generally underdetermined. We used the extra degrees of freedom to impose additional moment vanishing conditions on some of the basis elements, following the construction described by Alpert in [3][section 1.1], but this freedom could also be used to impose other conditions. For example, consider the Alpert space $L^2_{[0,1),2,2}(\mathbb{R})$ which has dimension 2 and contains piecewise-linear functions which are orthogonal to linear functions. Rather than imposing quadratic orthogonality on one basis element, we could instead use Alpert's projection technique to remove the discontinuity from the interior of one of the basis elements. Further investigation in this area would lie more in the realm of application than in the immediate scope of our project; we include it here for the sake of completeness.

Chapter 4, by contrast, should be considered only a first step in the investigation of measure-theoretic stability. We showed in section 4.3 that any Alpert basis in a one dimensional doubling measure is stable under small translations. However we also showed that the technique we used is insufficient to prove stability even for Haar wavelets in $\mathbb{R}^2$. Given that the Haar basis is known to be stable in $L^2(\mathbb{R}^d)$, shown through other means by Wilson in [19], we conjecture that our stability result for Alpert bases remains true in higher dimensions. Also, while we were interested only in Alpert bases, the material in section 4.2 applies equally to any orthonormal basis for $L^2(\mu)$ and to any measure μ on $\mathbb{R}^n$. It is possible that interesting results could be found by applying this technique in other contexts.

We also provided some initial investigation into the question of whether the doubling condition is necessary for stability. We showed that a measure on $\mathbb{R}$ cannot contain any point masses without causing some Haar bases to become unstable under small translations, however there are non-doubling measures which do not have this feature. Given that the doubling condition is closely related to small translations of dyadic intervals it seems a natural conjecture that the two conditions are equivalent, but we have been so far unable to locate a proof.

Appendix A

Completeness of Haar Wavelets

In this appendix we show that the Haar functions are complete in $L^2(\mathbb{R})$, and consequently that they form a basis. This completes the proof of theorem 1 from section 2.3. Recall that we have

$$h^I(x) = \frac{1}{\sqrt{l(I)}} \left(\mathbf{1}_{I_l}(x) - \mathbf{1}_{I_r}(x)\right)$$

and that we have already showed

1. $\{h^I\}_{I \in D}$ is an orthonormal set in $L^2(\mathbb{R})$.
2. Projection on to $\{h^I\}_{I \in D}$ satisfies a telescoping property: for any $f \in L^2(\mathbb{R})$ and integers $m < n$,

$$\sum_{I \in D \colon 2^{m+1} \leq l(I) \leq 2^n} \left\langle f, h^I \right\rangle h^I(x) = \sum_{I \in D \colon l(I) = 2^m} \left\langle f, \frac{1}{2^m} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) - \sum_{I \in D \colon l(I) = 2^n} \left\langle f, \frac{1}{2^n} \mathbf{1}_I \right\rangle \mathbf{1}_I(x).$$

3. Each h^I satisfies a moment vanishing condition:

$$\int_I h^I(x)\, dx = 0.$$

Before we proceed further, we define the *expectation* functions

$$\mathbb{E}_k f(x) = \sum_{I \in D \colon l(I) = 2^k} \left\langle f, \frac{1}{2^k} \mathbf{1}_I \right\rangle \mathbf{1}_I(x), \qquad x \in \mathbb{R}, k \in \mathbb{Z}.$$

This allows us to restate the telescoping property in the more compact form:

$$\sum_{I \in D:\, 2^{m+1} \leq l(I) \leq 2^n} \left\langle f, h^I \right\rangle h^I(x) = \mathbb{E}_m f(x) - \mathbb{E}_n f(x), \quad m < n.$$

Remark 6. *An observant reader will have noticed that we are reusing the* $\mathbb{E}$ *symbol, which also appeared in the definition of weighted Alpert wavelets. In that setting* $\mathbb{E}^{\mu}_{Q,U}$ *was a projection onto the space spanned by the functions in* U *on a single dyadic cube* Q*. In this setting,* $\mathbb{E}_k$ *is instead projection onto all dyadic intervals of length* 2^k*, rather than just a single interval. The notation could be made consistent by replacing each* $\mathbb{E}_k$ *here with* $\sum_I \mathbb{E}_I, k$*, but it would render this section much less pleasant to read.*

Lemma 14. *For every* $f \in L^2(\mathbb{R})$ *and every* $k \in \mathbb{Z}$*,* $\|\mathbb{E}_k f\|_2 \leq \|f\|_2$*.*

Proof. Fix $x \in \mathbb{R}$ and let $I \in D$ be the unique dyadic interval with $l(I) = 2^k$ and $x \in I$. Then

$$\begin{aligned}
|\mathbb{E}_k f(x)|^2 &= \left| \left\langle f, \frac{1}{2^k} \mathbf{1}_I \right\rangle \mathbf{1}_I(x) \right|^2 \\
&= \frac{1}{2^{2k}} \left| \int_I f(t)\, dt \right|^2 \\
&\leq \frac{1}{2^{2k}} \int_I |f(t)|^2\, dt \cdot \int_I 1\, dt \\
&= \frac{1}{2^{2k}} \int_I |f(t)|^2\, dt \cdot 2^k \\
&= \frac{1}{2^k} \int_I |f(t)|^2\, dt
\end{aligned}$$

by the Cauchy-Schwarz inequality. Since $\mathbb{E}_k f$ is constant on dyadic intervals, we get

$$\begin{aligned}
\int_{\mathbb{R}} |\mathbb{E}_k f(x)|^2\, dx &= \sum_{J \in D:\, l(J) = 2^k} \int_J |\mathbb{E}_k f(x)|^2\, dx \\
&\leq \sum_{J \in D:\, l(J) = 2^k} 2^k \cdot \frac{1}{2^k} \int_J |f(t)|^2\, dt \\
&= \sum_{J \in D:\, l(J) = 2^k} \int_J |f(t)|^2\, dt \\
&= \int_{\mathbb{R}} |f(t)|^2\, dt \\
&= \|f\|_2^2.
\end{aligned}$$

Therefore $\|\mathbb{E}_k f\|_2 \leq \|f\|_2$ as desired. □

Next, we observe that for $f \in L^2(\mathbb{R})$ and $k \in \mathbb{Z}$ we have

$$\mathbb{E}_k f(x) - \mathbb{E}_{k+1} f(x) = \sum_{I \in D \colon l(I) = 2^k} \langle f, h^I \rangle h^I(x).$$

Then by the telescoping proprerty of the Haar functions we have

$$\sum_{I \in D} \langle f, h^I \rangle h^I(x) = \lim_{M \to -\infty} \mathbb{E}_M f(x) - \lim_{N \to \infty} \mathbb{E}_N f(x).$$

If we can show that this difference of limits converges to $f(x)$ with respect to the L^2-norm, we will have proved completeness of the Haar functions.

To proceed, we will use the well-known but somewhat technical result that continuous, compactly-supported functions are dense in $L^2(\mathbb{R})$ (see [8, Proposition 7.9]). Let $f \in L^2(\mathbb{R})$ and $\epsilon > 0$; using this density, decompose f as $f = g_1 + g_2$ where $g_1 \in L^2(\mathbb{R})$ is continuous and supported inside a compact interval, and $g_2 \in L^2(\mathbb{R})$ has norm $\|g_2\|_2 < \frac{\epsilon}{2}$. Since g_1 is continuous we have $\mathbb{E}_k g_1(x) \to g_1(x)$ as $k \to -\infty$ for every $x \in \mathbb{R}$, and since g_1 is compactly supported we can conclude $\|g_1 - \mathbb{E}_k g_1\|_2 \to 0$ as $k \to -\infty$. We also have

$$\left| \left\langle f, \frac{1}{l(I)} \mathbf{1}_I \right\rangle \right| = \left| \frac{1}{l(I)} \int_I f(x)\, dx \right| \le \left(\frac{1}{l(I)} \int_I |f(x)|^2\, dx \right)^{\frac{1}{2}} = \frac{1}{\sqrt{l(I)}} \|f\|_2,$$

so $\mathbb{E}_k f(x) \to 0$ as $k \to \infty$ for every $x \in \mathbb{R}$.

Finally by lemma 14 and several applications of the triangle inequality we have

$$\begin{aligned}
\left\| f - \sum_{I \in D} \langle f, h^I \rangle h^I \right\|_2 &= \left\| f - \left(\lim_{M \to -\infty} \mathbb{E}_M f - \lim_{N \to \infty} \mathbb{E}_N f \right) \right\|_2 \\
&\leq \left\| f - \lim_{M \to -\infty} \mathbb{E}_M f \right\|_2 + \left\| \lim_{N \to \infty} \mathbb{E}_N f \right\|_2 \\
&\leq \left\| g_1 + g_2 - \lim_{M \to -\infty} \mathbb{E}_M (g_1 + g_2) \right\|_2 + 0 \\
&\leq \left\| g_1 - \lim_{M \to -\infty} \mathbb{E}_M g_1 \right\|_2 + \left\| g_2 - \lim_{M \to -\infty} \mathbb{E}_M g_2 \right\|_2 \\
&\leq 0 + \|g_2\|_2 + \|g_2\|_2 \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}$$

Therefore $f = \sum_{I \in D} \langle f, h^I \rangle h^I$ in $L^2(\mathbb{R})$, and we conclude that the Haar functions are complete in $L^2(\mathbb{R})$.

Appendix B

Buchberger's Algorithm

Here we give one example of an algorithm for computing Gröbner bases. The running time of such algorithms can vary heavily depending on the choice of monomial order and on various choices made during computation; practical implementations will use algorithms more complex than the one presented here. Recall that a Gröbner basis depends on the choice of monomial order M.

Theorem 12. *Let $G = \{g_1, \ldots, g_k\} \subset \mathbb{R}[\mathbf{x}]$ be a finite set of polynomials. Then every $f \in \mathbb{R}[\mathbf{x}]$ can be expressed as $f = q_1 g_1 + \cdots + q_k g_k + r$ where $q_i, r \in \mathbb{R}[\mathbf{x}]$ and either $r = 0$ or r is a linear combination of monomials not divisible by any $LT(g_i)$.*

This theorem follows immediately from the multivariate division algorithm, which generates appropriate choices of q_i, r given f, G:

1. Let $q_1 = \cdots = q_k = r = 0$.

2. If $LT(g_i)$ divides $LT(f)$ for some $i \in \{1, \ldots, k\}$: replace f by $f - \frac{LT(f)}{LT(g_i)} g_i$, add $\frac{LT(f)}{LT(g_i)}$ to q_i, and restart step 2. If $LT(g_i)$ does not divide $LT(f)$ for any i, continue to step 3.

3. Add $LT(f)$ to r, then replace f by $f - LT(f)$. If now $f = 0$, stop. Otherwise return to step 2.

Since the leading term of f is always reduced (with respect to M) in step 2, this algorithm is guaranteed to terminate. We note that this algorithm does not produce a unique decomposition, as the output can depend on choices of i in step 2 of the algorithm. Despite this lack of uniqueness, the divisibility condition on r preserves the intuitive notion that the remainder ought to be "smaller" in some sense than the divisors.

Definition 22 (S-polynomial)**.** *Given two polynomials $f, g \in \mathbb{R}[\mathbf{x}]$, let $a = lcm(LT(f), LT(g))$. Define the S-polynomial of f and g to be $S(f, g) := \frac{a}{LT(f)}f - \frac{a}{LT(g)}g$.*

By construction the leading terms of $\frac{a}{LT(f)}f$ and $\frac{a}{LT(g)}g$ are equal, so they cancel when computing S. This definition is motivated by the following:

Theorem 13 (Buchberger's Criterion)**.** *Let $G = \{f_1, ..., f_2\}$ be a generating set of polynomials for I. Then G is a Gröbner basis for I if and only if every S-polynomial $S(f_i, f_j)$, $i \neq j$ yields a remainder of 0 when divided by the elements of G in some order.*

A proof of this result can be found in [7, p. 324]. If we take G to be an arbitrary generating set for I, we then have Buchberger's algorithm:

1. Choose two polynomials $f_i, f_j \in G$, $i \neq j$ and compute the S-polynomial $S(f_i, f_j)$.
2. Divide $S(f_i, f_j)$ by G. If the resulting remainder r is non-zero, add r to G.
3. Repeat steps 1 and 2 until all possible pairs have been considered, including all polynomials added in step 2.
4. Output G.

The output set G satisfies the condition in Theorem 13. In step 2 the addition of new elements to G strictly increases $LT(G)$ and as $\mathbb{R}[\mathbf{x}]$ is Noetherian this process must eventually terminate. Loosely speaking, the process of computing remainders of S-polynomials within G produces any "hidden" polynomials in I which are linear combinations of the generators.

Bibliography

[1] Alexis, M., Sawyer, E., and Uriarte-Tuero, I. Tops of dyadic grids and $T1$ theorems. https://arxiv.org/abs/2201.02897, 2022.

[2] Alexis, M., Sawyer, E. T., and Uriarte-Tuero, I. A $T1$ theorem for general smooth Calderón-Zygmund operators with doubling weights, and optimal cancellation conditions, II. *J. Funct. Anal. 285*, 11 (2023), Paper No. 110139, 52.

[3] Alpert, B. K. A class of bases in L^2 for the sparse representation of integral operators. *SIAM J. Math. Anal. 24*, 1 (1993), 246–262.

[4] Christensen, O. *An Introduction to Frames and Riesz Bases.* Applied and Numerical Harmonic Analysis. Springer International Publishing, 2016.

[5] Daubechies, I. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math. 41*, 7 (1988), 909–996.

[6] David, G., and Journe, J.-L. A boundedness criterion for generalized calderon-zygmund operators. *Annals of Mathematics 120*, 2 (1984), 371–397.

[7] Dummit, D., and Foote, R. *Abstract Algebra*, 3 ed. Wiley, 2003.

[8] Folland, G. *Real Analysis: Modern Techniques and Their Applications.* Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2013.

[9] Garnett, J., Killip, R., and Schul, R. A doubling measure on $\mathbb{R}^d$ can charge a rectifiable curve. *Proc. Amer. Math. Soc. 138*, 5 (2010), 1673–1679.

[10] Haar, A. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen 69*, 1 (1910), 331–371.

[11] Kovalev, L., Maldonado, D., and Wu, J.-M. Doubling measures, monotonicity, and quasiconformality. *Math. Z. 257*, 3 (2007), 525–545.

[12] Nazarov, F., Treil, S., and Volberg, A. Two weight estimate for the Hilbert transform and corona decomposition for non-doubling measures. https://arxiv.org/abs/1003.1596, 2010.

[13] Pereyra, M., and Ward, L. *Harmonic Analysis: From Fourier to Wavelets*. IAS/Park city mathematical subseries. American Mathematical Society, 2012.

[14] Rahm, R., Sawyer, E. T., and Wick, B. D. Weighted Alpert wavelets. *J. Fourier Anal. Appl. 27*, 1 (2021), Paper No. 1, 41.

[15] Sawyer, E. T. $T1$ testing implies Tp polynomial testing: optimal cancellation conditions for CZO's. https://arxiv.org/abs/1907.10734, 2019.

[16] Sawyer, E. T. A probabilistic analogue of the fourier extension conjecture. https://arxiv.org/abs/2311.03145, 2024.

[17] Stanković, R. S., and Falkowski, B. J. The Haar wavelet transform: its status and achievements. *Computers & Electrical Engineering 29*, 1 (2003), 25–44.

[18] Stein, E., and Shakarchi, R. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2009.

[19] Wilson, M. Bounded variation, convexity, and almost-orthogonality. In *Harmonic analysis, partial differential equations and applications*, Appl. Numer. Harmon. Anal. Birkhäuser/Springer, Cham, 2017, pp. 275–301.