ABSTRACT

Title of dissertation: MULTISCALE ANALYSIS AND DIFFUSION SEMIGROUPS WITH APPLICATIONS

Karamatou Yacoubou Djima
Doctor of Philosophy, 2015

Dissertation directed by: Professor Wojciech Czaja
Mathematics Department

Multiscale (or multiresolution) analysis is used to represent signals or functions at increasingly high resolution. In this thesis, we develop multiresolution representations based on frames, which are overcomplete sets of vectors or functions that span an inner product space.

First, we explore composite frames, which generalize certain representations capable of capturing directionality in data. We show that we can obtain composite frames for $L^2(\mathbb{R}^n)$ given two main ingredients: 1) dilation operators based on matrices from admissible subgroups $\mathbf{G}_A$ and $\mathbf{G}$, and 2) a generating function that is refinable with respect to $\mathbf{G}_A$ and $\mathbf{G}$.

We also construct frame multiresolution analyses (MRA) for L^2-functions of spaces of homogeneous type. In this instance, dilations are represented by operators that come from the discretization of a compact symmetric diffusion semigroup. The eigenvectors shared by elements of the compact symmetric diffusion semigroup can be used to define an orthonormal MRA for L^2. We introduce several frame systems

that generate an equivalent MRA, notably composite diffusion frames, which are built with the composition of two "similar" compact symmetric diffusion semigroups.

The last part of this thesis is an application of Laplacian Eigenmaps (LE) to a biomedical problem: Age-Related Macular Degeneration. LE, a tool in the family of diffusion methods, uses similarities at local scales to provide global analysis of data sets. We propose a novel approach with two steps. First, we apply LE to retinal images, provided by the National Institute of Health, for feature enhancement and dimensionality reduction. Then, using an original Vectorized Matched Filtering technique, we detect retinal anomalies in eigenimages produced by the LE algorithm.

MULTISCALE ANALYSIS AND DIFFUSION SEMIGROUPS WITH APPLICATIONS

by

Karamatou Yacoubou Djima

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Wojciech Czaja, Chair
Professor John J. Benedetto
Professor Kasso Okoudjou
Professor Radu Balan
Professor Ramani Duraiswami, Dean's Representative

Dedication

Au nom de Dieu, le plus Clément et le plus Miséricordieux

Acknowledgments

First and foremost, I would like to thank my advisor, Wojciech Czaja, for his incredible support throughout my graduate studies. From my second semester on, Wojciech has taught me what it means to be an applied mathematician. Wojciech has allowed me to explore different topics and become a more cultivated applied harmonic analyst, an endeavor for which I could not have hoped for a better model than him. Wojciech has always provided me with the right amount of guidance and laissez-faire in every step of my research. I am deeply appreciative for Wojciech's patience and flexibility whenever I expressed doubts about which direction to follow. Wojciech has also insisted that I keep a balance in my life and never lose sight of what I value most. I hope to carry this wisdom in my future. I am most grateful for the support that Wojciech has offered to me during some of the earliest, most challenging years of my stay in the math department. Without all this, I could not have succeeded.

I am grateful to other faculty members in the Norbert Wiener Center. John Benedetto has shown an unwavering trust that I would succeed whenever I doubted myself, and provided me with the kindness and wisdom that I needed in those times. I will forever cherish my conversations with John, as well as his lectures, including entertaining anecdotes on mathematicians. I also thank Kasso Okoudjou for his friendliness throughout the years and invaluable help in parts of my thesis. I truly enjoyed our conversations on various topics and admire Kasso's dedication to mathematics. Radu Balan has been an inspiration through his inquisitive nature. Finally, I thank Patrick Fitzpatrick for being a paragon of excellence in teaching.

I owe thanks to Lawrence Washington for all his support at various stages, Ramani Duraiswami, my Dean's representative, for his kindness and for pointing to some clarifications needed in my thesis and Konstantina Trivisa for accepting me in AMSC, hence, giving me the opportunity to fulfill my dream.

I am thankful for the amazing administrative staff at the math department, in particular, Alverda McCoy, Rhyneta Gumbs and Bill Schildknecht. Without Alverda, administrative matters could have easily turned into insurmountable conundrums. Alverda has also always been so kind and encouraging! Our conversations were a source of renewed enthusiasm for me. Rhyneta has always been helpful and done anything in her capability to solve any issue I (or others) could have. Rhyneta has also shown congeniality, care, constant support; I am honored to call her a friend. I thank Bill for being so accommodating when I needed it most.

I cannot imagine these PhD years without the friendship and mentoring of many wonderful people. First of all, I feel so much gratitude toward to the "math ladies" in my life! Our relationship is a gift I have cherished every day, whether it has taken the form of passionate conversations, dancing together, comforting hand holding... I thank Marie for her companionship (these fun and uplifting gchat sessions!), sincere care and reminders to remain sane and grateful; Clare, for her inspiring work ethic and morals, her immense generosity and for making me feel important; Jennie, for being kind, loving and engaging; Lucia, for her thoughtfulness and admirable intellectual honesty; Hana, for being an example of quiet determination. I owe much of the courage to persevere and believe in myself to the amazing Shelby Wilson, my mentor, my model, one of my biggest cheerleaders. I am also lucky to have in my circle of

friends: Hisham, with his strange ability to make me laugh and mad like no one else can, Alex, with his affectionate nature and stimulating intellectual curiosity, Kevin, with his contagious optimism, Paul, with whom I shared trials and victories and Garrett, with his considerate personality and great jokes. I am appreciative for many others in my journey. My awesome NWC colleagues: Tim, Wei-Hsuan, Ariel, Travis, the Matt's, Chae, Ben, Julia. My office neighbors: Ran, Jonathan, Richard and Srimathy. There are also the fantastic EDGE program (Ulrica, Amy, etc.), Deena, Sika, Cathy, Patrick, the 1305 office, the Ultimate group and so many more!

I thank all those who were part of my education. I especially thank my undergraduate advisor, Andrew Poje, for believing in me, teaching me critical material and guiding me toward a PhD in applied math. I promised that someday I would make Drew proud; I hope that I achieved this, at least partially.

Last, but certainly not least, I want to thank my family. As in the poem, "Que pourrais-je vous dire que vous ne sachiez déjà?" Our close-knit bound is, by far, the most precious gift I have in this world. You are my core, my inspiration, my reason for persevering. You manage to uplift me in all circumstances through your sense of humor and deep care. To Papa et Maman, who know nothing but courage and dedication toward me and my siblings: your faith in me has carried me since my birth and continues to be the only constant in my life. I could never thank you enough. To my siblings Abiba and Ismael, for being my sounding board, for having so much patience, generosity and kindness toward me: you are amazing human beings; I will always strive to be more like you and to deserve having you in my life.

To anyone who, close or far, brought any contribution to this work. Thank you.

Table of Contents

List of Notations

$\mathbb{R}$	the field of real numbers
$\mathbb{N}$	the set of natural numbers
$\mathbb{C}$	the field of complex numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\hat{\mathbb{R}}^n$	n-dimensional frequency domain
$\mathbb{Z}^n$	the set of n-dimensional integers
$\#\|A\|$	the cardinality of the set A
$\operatorname{supp}(f)$	the support of f
$\|A\|$	the volume/area of the set A
∂U	the boundary of the set U
$B_R(x)$	the ball of radius R centered at x
$a.e.$	almost everywhere
$L^p(X)$	the set of functions f such that $\|f\|^p$ is integrable on X with respect to the Lebesgue measure
$\hat{f}$, $\mathscr{F}[f]$	the Fourier transform of the function f
T^*	the adjoint of the operator T
M_y	the modulation by y
D_A	the dilation with respect to the matrix A
$GL_n(S)$	the general linear group over S
$SL_n(S)$	the special linear group over S
$\forall$	for all
V_j, V_j^{ε}	approximation spaces
$\{S^t\}_{t\geq 0}$	symmetric diffusion semigroup
$\sigma(S)$	the spectrum of S

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The availability of high dimensional data has soared in the past twenty years. In all areas of science, mathematics and even in popular culture, the ubiquity of expressions such as "big data" or, more dramatically, "data deluge", demonstrates the scientific appeal of this phenomenon. Our interest in constructing efficient, sparse, representation systems for such large data is driven both by pure mathematical pursuit and the reality of technology today. Indeed, although the sustained increase of computational power facilitates the acquisition of immense amounts of data, most of that data could be meaningless without the proper analysis. Therefore, the goal in many areas of mathematical sciences is to find techniques to represent data efficiently for clustering, prediction, visualization, etc. In designing these techniques, one or more of the following properties can often be exploited:

- **Intrinsic low-dimensionality.** The data is embedded in high dimensional space, but it actually lies on a significantly lower dimensional space due to physical or statistical constraints.
- **Varying behavior at different scales.** An example of this could be a physical process that obeys different laws at different scales. In this case, Multiscale (or

Multiresolution) Analysis, i.e., data representation at each scale, can reveal and efficiently capture changing behaviors of the systems at different scales.

In this thesis, the systems that we study, Composite Frames (or Frame $\mathbf{G}_A\mathbf{G}$-Multiresolution Analysis), Diffusion (Wavelet) Frames and Composite Diffusion Frames, give an efficient multiresolution representation for data possessing one or both of the properties above. In Chapter 2 and 3, we lay out some theoretical foundations for these methods. In Chapter 4, we present the application of Laplacian Eigenmaps (LE), a member of the family of diffusion methods, in combination with a matched filtering-based algorithm, Vectorized Matched-Filtering (VMF), in a retinal imaging problem.

Although dealing with large data sets is a concern typical to our era, the search of efficient representation systems is a problem as old as age. From its origin with Fourier Analysis ([49]) to Wavelet analysis ([28, 31]) and its successors, Harmonic Analysis has engendered many successful methods that respond to the challenge of efficient representation by decomposing signals or data into basic constituents. The classical example of time-frequency methods, Fourier Analysis, uses orthogonal eigenfunctions of the Laplace operator on subsets of $\mathbb{R}^n$ as a basis for square-integrable functions or signals defined on subsets of $\mathbb{R}^n$. For many problems, Fourier analysis gives good global approximations and there exist fast transforms for efficient computations. However, Fourier representation elements are not localized in space, which can create significant approximation errors in certain applications [33]. In Wavelet Analysis, square-integrable functions defined on subsets of $\mathbb{R}^n$ are approximated using wavelets, which are orthonormal sets of functions localized both in space and

frequency. There also exist fast transforms for wavelet analysis and, moreover, functions can be represented at different scales (or resolutions), which allows for fast pyramidal schemes in numerical computations [33, 78]. Based on the success of these harmonic analysis tools on $\mathbb{R}^n$, in recent years, there has been a great interest in establishing Fourier- and wavelet-type analysis for non-Euclidean domains. In 2006, Coifman and Lafon constructed Diffusion Maps, a Fourier basis for square-integrable functions of domains such as manifolds and graphs [26], and in the same year, Coifman and Maggioni introduced Diffusion Wavelets, a wavelet basis for the same type of functions and domains [27].

Briefly, Diffusion Wavelets are an efficient wavelet multiscale analysis for functions on domains such as graphs and manifolds for which the notion of scale is not associated with any "natural" operation such as dilation on Euclidean spaces. Diffusion wavelets (and the accompanying wavelet packets [13]) are useful to represent many functions on graphs and manifolds with good accuracy. However, their construction is associated with a high computational cost due to an underlying orthogonalization process [27]. The present dissertation expands the theory of diffusion methods in ways that we describe more precisely in the thesis contribution, Section 1.3.

In the next few sections of this chapter, we present preliminary concepts and introduce the notation and definitions that will be used throughout the thesis.

1.2 Preliminaries

1.2.1 L^p-Spaces & Fourier Transform

We have already been using $\mathbb{R}^n$ to represent Euclidean spaces; let us denote by $\mathbb{N}$, $\mathbb{C}$ and $\mathbb{Z}^n$, the set of natural numbers, complex numbers and the n-dimensional integer lattice, respectively.

Let x be a column vector representing points in $\mathbb{R}^n$ and ω, a row vector representing points in the frequency domain $\hat{\mathbb{R}}^n$. Suppose X is an open subset of $\mathbb{R}^n$ or $\mathbb{R}^n$ itself. For $1 \leq p < \infty$, the Banach spaces $L^p(X)$ contain complex-valued functions f for which $|f|^p$ is integrable on X with respect to the Lebesgue measure, i.e.,

$$L^p(X) := \left\{ f : X \longrightarrow \mathbb{C} : f \text{ is measurable and } \int_X |f(x)|^p \mathrm{d}x < \infty \right\}.$$

The norm on $L^p(X)$ is

$$\|f\|_p = \|f\|_{L^p(X)} = \left(\int_X |f(x)|^p \mathrm{d}x \right)^{1/p}.$$

For $p = \infty$, the **essential supremum** of a function f on X is given by

$$\|f\|_{L^\infty(X)} = \operatorname*{ess\,sup}_{x \in X} f = \inf_{x \in X} \{\lambda \in \mathbb{R} : f(x) \leq \lambda \text{ a.e.}\}.$$

We are particularly interested in the space $L^2(X)$ of complex-valued, square integrable functions on X, with respect to the Lebesgue measure:

$$L^2(X) := \left\{ f : X \longrightarrow \mathbb{C} : f \text{ is measurable and } \int_X |f(x)|^2 dx < \infty \right\},$$

equipped with the norm

$$\|f\|_2 = \|f\|_{L^2(X)} = \left(\int_X |f(x)|^2 \mathrm{d}x \right)^{1/2}.$$

Note that, in agreement with standard notation, we sometimes write $\|f\|$ to denote $\|f\|_2$. The space $L^2(X)$ is a Hilbert space and thus has an inner product given by

$$\langle f,\, g \rangle = \int_X f(x)\overline{g(x)}\mathrm{d}x, \quad f,\, g \in L^2(X).$$

To compute estimates for certain quantities associated with functions in $L^2(X)$, we will often use the Cauchy-Schwartz inequality, which states that

$$\int_X |f(x)g(x)|\mathrm{d}x \leq \left(\int_X |f(x)|^2 \right)^{1/2} \left(\int_X |g(x)|^2 \right)^{1/2} \quad \text{for all } f,\, g \in L^2(X).$$

The discrete analogue of $L^2(\mathbb{R}^n)$ is $\ell^2(\mathrm{K})$, the space of square summable scalar sequences with respect to a countable index K:

$$\ell^2(\mathrm{K}) := \left\{ \{x_k\}_{k\in \mathrm{K}} \subseteq \mathbb{C},\, \mathrm{K} \text{ is countable} : \|f\|^2_{\ell^2(\mathrm{K})} = \sum_{k\in \mathrm{K}} |x_k|^2 < \infty \right\}.$$

The space $\ell^2(\mathrm{K})$ is also a Hilbert space with respect to the inner product

$$\langle \{x_k\},\, \{y_k\} \rangle = \sum_{k\in \mathrm{K}} x_k \overline{y_k},$$

where $\{x_k\}_{k\in \mathrm{K}},\, \{y_k\}_{k\in \mathrm{K}} \subset \ell^2(\mathrm{K})$.

The Cauchy-Schwartz inequality on $L^2(\mathrm{K})$ is given by

$$\left| \sum_{k\in \mathrm{K}} x_k \overline{y_k} \right|^2 \leq \sum_{k\in \mathrm{K}} |x_k|^2 \sum_{k\in \mathrm{K}} |y_k|^2,\ \{x_k\}_{k\in \mathrm{K}},\, \{y_k\}_{k\in \mathrm{K}} \subset \ell^2(\mathrm{K}).$$

Finally, we denote by $\mathbb{T}^n$ the n-dimensional torus $\mathbb{R}^n/\mathbb{Z}^n \simeq [0\,1]^n$. The space of

measurable $\mathbb{Z}^n$ periodic function f such that

$$\|f\|^2_{L^2(\mathbb{T}^n)} := \int_{\mathbb{T}^n} |f(x)|^2 \mathrm{d}x < \infty$$

is denoted by $L^2(\mathbb{T}^n)$.

Next, we define the Fourier transform of functions in $L^2(\mathbb{R}^n)$. Note that the Fourier transform is usually defined for functions in $L^1(\mathbb{R}^n)$, but because we wish to use formulas such as Plancherel's equation without additional assumptions, we adopt the following definition.

Definition 1.1. The **Fourier transform** of $\mathscr{F} : L^2(\mathbb{R}^n) \longrightarrow L^2(\hat{\mathbb{R}}^n)$ of a function $f \in L^2(\mathbb{R}^n)$ is given by

$$\hat{f}(\omega) := \mathscr{F}[f](\omega) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i \omega x} \, \mathrm{d}x, \; \omega \in \hat{\mathbb{R}}^n. \tag{1.1}$$

For all $f, g \in L^2(\mathbb{R}^n)$, we have **Plancherel's equation**

$$\langle f, g \rangle = \left\langle \hat{f}, \hat{g} \right\rangle, \text{ and } \|f\|_{L^2(\mathbb{R}^n)} = \|\hat{f}\|_{L^2\left(\hat{\mathbb{R}}^n\right)}. \tag{1.2}$$

1.2.2 Frames

Although frames only started to garner popularity under the influence of Daubechies, Grossman and Meyer in the 1980's [34], they were introduced in 1952 by Duffin and Schaeffer [42], and even before that, they were a recurrent topic in the mathematical literature. Under the influence of Benedetto, Casazza, Christensen, and a few others, the importance and usefulness of frames is now fully established, and they engender significant interest both theoretically and in practice, particularly in signal analysis applications [1, 3, 7, 8, 17, 18, 23].

In a nutshell, frames can be described as a redundant (or over-complete) set of vectors or functions that spans an inner product space $\mathcal{V}$. Frames are often compared to **orthonormal bases** (ONB's), which are also a spanning set but have the added requirements that the elements are 1) linearly independent, 2) orthogonal, and 3) have norm 1. The main consequence of these differences between frames and ONB's is that vectors or functions in $\mathcal{V}$ have a unique representation in ONB's, unlike in frames, because of redundancy. Uniqueness of representation is a key property that makes reconstructing a signal from its ONB decomposition computationally tractable and stable. Therefore, frames may lack an important feature needed for signal reconstruction. However, despite this drawback, frames present attractive features. One is flexibility. Indeed, since frames can be constructed without independence and orthogonality restrictions, they allow for varied characteristics that can be custom-made for the application of interest. Moreover, non-uniqueness of representation can actually lead to a more robust representation of vectors or functions in $\mathcal{V}$ during certain processes [60]. For example, suppose that you would like to send a signal across some communication system using the coefficients which represent the signal in terms of an ONB or in terms of a frame system. As mentioned above, each element of the ONB will be associated with a single coefficient. If one coefficient is lost in the transmission, its information content cannot be recovered. With frames, however, if one piece is missing, because of redundancy, the information can be recovered from the remaining pieces [60].

More importantly, a goal of our work is to perform analysis of the data in a reduced dimension space, so there may be no need to reconstruct the original data.

Therefore, we can fully take advantage of the desirable features of frames without suffering the main inconvenience. Note also that, in some well-established cases, a signal reconstruction from its frame decomposition is just as computationally feasible as for ONB's.

Now, we give the formal definition of frames.

Definition 1.2. A countable family of elements $\{f_k\}_{k=1}^{\infty}$ in a (separable) Hilbert space $\mathcal{H}$ is a **frame** for $\mathcal{H}$ if for each $f \in \mathcal{H}$ there exist constants C_L, $C_U > 0$ such that

$$C_L\|f\|^2 \leq \sum_{k=1}^{\infty} |\langle f,\, f_k\rangle|^2 \leq C_U\|f\|^2. \tag{1.3}$$

The constants C_L and C_U are called the frame bounds: C_L is the **lower frame bound** and C_U is the **upper frame bound**, and they are optimal if C_L is maximal and C_U is minimal. When a countable family of elements in a Hilbert space satisfies the upper bound condition, we say this family is a **Bessel system**.

Different types of frames can be defined in terms of the value of frame bounds. We give a few types as follows:

Definition 1.3. a) A frame is **tight** if $A = B$.

b) A frame is a **Parseval frame** if it is tight with $A = 1$.

c) A frame is a **finite unit-norm tight frame** (FUNTF) if it is tight and each frame element has norm one.

Example 1.4. *An orthonormal basis satisfies the frame definition with $C_L = C_U = 1$, so an ONB is a Parseval frame, but also a FUNTF since each element is unit norm.*

This shows that frames are a generalization of orthonormal bases.

Next, we study the operators associated with frames. Recall that the following definition an adjoint:

Definition 1.5. The **adjoint** of a bounded linear operator $T : \mathcal{H} \longrightarrow \mathcal{K}$, where $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces, is the unique operator $T^* : \mathcal{K} \longrightarrow \mathcal{H}$ satisfying the inner production equality

$$\langle Tx,\, y\rangle_{\mathcal{K}} = \langle x,\, T^*y\rangle_{\mathcal{H}}\,, \text{ for all } x \in \mathcal{H},\, y \in \mathcal{K}.$$

When $T = T^*$, we say that T is **self-adjoint**.

Note that definition implies that $\|T\| = \|T^*\|$.

Definition 1.6. Let $\{f_k\}_{k=1}^{\infty}$ be a frame in a Hilbert space $\mathcal{H}$ and let $f \in \mathcal{H}$.

a) The **synthesis operator** T is given by

$$T : \ell^2(\mathbb{N}) \longrightarrow \mathcal{H}, \quad T\{c_k\}_{k=1}^{\infty} = \sum_{k=1}^{\infty} c_k f_k. \tag{1.4}$$

a) The **analysis operator** is given by the adjoint of T,

$$T^* : \mathcal{H} \longrightarrow L^2(\mathbb{N})\,, \quad T^*f = \{\langle f,\, f_k\rangle\}_{k=1}^{\infty}\,. \tag{1.5}$$

c) The **frame operator** is obtained by composing T and T^*:

$$S : \mathcal{H} \longrightarrow \mathcal{H}, \quad Sf = TT^* \sum_{k=1}^{\infty} \langle f,\, f_k\rangle\, f_k. \tag{1.6}$$

The operator S is bounded, invertible, self-adjoint, and positive.

The operators associated with frames govern the interaction of frame elements with

other constituents of the Hilbert space. The following result concerns the characterization of a frame system via the synthesis operator T.

Proposition 1.7. *A sequence $\{f_k\}_{k=1}^{\infty}$ in $\mathcal{H}$ is a frame for $\mathcal{H}$ if and only if*

$$T : \{c_k\}_{k=1}^{\infty} \longrightarrow \sum_{k=1}^{\infty} c_k f_k$$

is a well-defined mapping from $\ell^2(\mathbb{N})$ onto $\mathcal{H}$.

Next, we look at properties of frames under the action of operators other than those naturally associated with them. The following results will come to play in Chapters 2 and 3, when we will have families of frames acted on by certain **"dilation"** operators. The meaning of dilation will vary from one chapter to the other, but, as we will discuss later, we can relate these dilation operators by some common properties of their action.

First, we give the formula for the frame operator of the frame obtained by applying an operator to a finite frame. This standard result can be found in [19].

Proposition 1.8. *Let $\{f_k\}_{k=1}^{N}$, $N \in \mathbb{N}$, be a frame in a finite dimensional Hilbert space $\mathcal{H}$, with frame operator S. If U is an operator on $\mathcal{H}$, then the frame operator for $\{Uf_k\}_{k=1}^{N}$ is USU^*. If U is invertible, then $\{Uf_k\}_{k=1}^{N}$ also constitutes a frame for $\mathcal{H}$.*

For the next result, first recall the definition of a unitary operator:

Definition 1.9. An operator $U : \mathcal{H} \longrightarrow \mathcal{H}$ is **unitary** if, for $x, y \in \mathcal{H}$,

$$\langle Ux, Uy \rangle_{\mathcal{H}} = \langle x, y \rangle_{\mathcal{H}},$$

which implies $U^* = U^{-1}$.

In [23], we have the following result:

Proposition 1.10. *Let $\{f_k\}_{k=1}^{\infty}$ be a frame in a Hilbert space $\mathcal{H}$ with frame bounds C_L, $C_U > 0$. If $U : \mathcal{H} \longrightarrow \mathcal{H}$ is a unitary operator, then $\{Uf_k\}_{k=1}^{\infty}$ is a frame with frame bounds C_L, C_U.*

Proposition 1.10 tells us that applying an unitary operator to a frame for $\mathcal{H}$ produces another frame for $\mathcal{H}$. This is very useful when we can find a unitary operator that gives better properties to our original frames, e.g., a more precise approximation of elements of $\mathcal{H}$.

Now, when the operator U applied to a frame is not unitary but has other desirable properties, the following proposition, also found in [23], will be helpful.

Proposition 1.11. *Let $\mathcal{H}$, $\mathcal{K}$ be Hilbert spaces, and suppose that $U : \mathcal{K} \longrightarrow \mathcal{H}$ is a bounded operator with closed range R_U. Then there exists a bounded operator $U^\dagger : \mathcal{H} \longrightarrow \mathcal{K}$, the pseudo-inverse of U, for which*

$$UU^\dagger f = f, \tag{1.7}$$

for all $f \in R_U$. Moreover, the following holds:

(i) The orthogonal projection of $\mathcal{H}$ onto R_U is given by $UU^\dagger$.

(ii) The orthogonal projection of $\mathcal{K}$ onto $R_{U^\dagger}$ is given by $U^\dagger U$.

(iii) U^* has closed range and $(U^*)^\dagger = (U^\dagger)^*$.

(iv) On R_U, the operator $U^\dagger$ is given explicitly by

$$U^\dagger = U^*(UU^*)^{-1}. \tag{1.8}$$

Proposition 1.11 leads to the following:

Corollary 1.12. *(i) Let $\mathcal{H}$, $\mathcal{K}$ be Hilbert spaces, and suppose that $U : \mathcal{K} \longrightarrow \mathcal{H}$ is a bounded operator with closed range R_U. Let $U^\dagger$ be defined as in* (1.7)*. If U is invertible, then*

$$U^\dagger = U^{-1}.$$

(ii) Suppose $U_1U_2 : \mathcal{K} \longrightarrow \mathcal{H}$ is a bounded, closed range operator with $U_2 : \mathcal{K} \longrightarrow \mathcal{K}' = R_{U_2}$, bounded, closed range and $U_1 : \mathcal{K}' \longrightarrow \mathcal{H}$, also bounded, closed range. Then

$$(U_1U_2)^\dagger = U_2^\dagger U_1^\dagger.$$

Proof. (i) We know $UU^\dagger f = f$ and $UU^{-1}f = f$. By uniqueness of inverse, it must be that $U^{-1} = U^\dagger$.

(ii) Using Proposition 1.11 (i), for $f \in \mathcal{H}$, we have

$$(U_1U_2)(U_1U_2)^\dagger f = f.$$

Now,

$$U_1U_2U_2^\dagger U_1^\dagger f \;\; = \;\; U_1(U_1^\dagger f),$$

where, we have used $U_2U_2^\dagger U_1^\dagger f = U_1^\dagger f$ since $U_1^\dagger f \in \mathcal{K}'$. Thus,

$$U_1U_2U_2^\dagger U_1^\dagger f = U_1(U_1^\dagger f) = f.$$

□

This result will be useful in Chapter 3, when we build a frame system for a Hilbert

space $\mathcal{H}$ using the discretization of a family of operators called compact symmetric diffusion semigroup. These operators are not unitary but are bounded and closed range, and they will allow us to obtain frames for approximation subspaces of the Hilbert space $\mathcal{H}$.

1.2.3 Wavelets and Multiresolution Analysis

Wavelets are another celebrated representation system brought to high interest in the 1980's. Since that time, they have given rise to a plethora of related representation systems that aim to give increasingly precise approximations of functions or signals, including performing tasks such as detecting singularities or denoising. We give an overview of important concepts in wavelet analysis by considering a basic example in $L^2(\mathbb{R})$.

Let $\psi \in L^2(\mathbb{R})$ and j, $k \in \mathbb{Z}$. Define

$$\psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k), \quad x \in \mathbb{R}. \tag{1.9}$$

If $\{\psi_{j,k}(x)\}_{j,\,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$, the function ψ is called a **wavelet** or **mother wavelet**.

Example 1.13. *A basic example of wavelet, the Haar wavelet, is defined by*

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \le x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Haar proved that, for this choice of ψ, the system $\{\psi_{j,k}(x)\}_{j,\,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$. For the proof, please, refer to [23, 33, 57].

The Haar wavelet has advantages such as simplicity, orthogonality and compact support. However, it is clear from the definition that it has poor differentiability properties, which can cause severe errors in certain approximations. To compensate for this, subsequent wavelets were designed with characteristics such as exponential decay or smoothness. The introduction of **multiresolution analysis** gave a systematic construction of wavelet orthonormal bases [77]. The definition is as follows.

Definition 1.14. A sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ together with a function ϕ is a **multiresolution analysis (MRA)** for $L^2(\mathbb{R})$ if the following hold:

(i) $\cdots V_{-1} \subset V_0 \subset V_1 \cdots$.

(ii) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$ and $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

(iii) $f \in V_j \iff f(2x) \in V_{j+1}$, $x \in \mathbb{R}$.

(iv) $f \in V_0 \Longrightarrow f(x-k) \in V_0$, for all $k \in \mathbb{Z}$, $x \in \mathbb{R}$.

(v) $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0.

The properties described in Definition 1.14 are very useful for approximations. For example, if we are looking for the approximation of a function $f \in L^2(\mathbb{R})$ in a certain space V_j and cannot find a satisfying one, we know, by (i), that the V_j's are nested, and this allows us to move to another approximation space $V_{j'}$, $j' \neq j$, via the simple scaling defined in (iii).

Starting with a MRA, one can define, for each $j \in \mathbb{Z}$, the space W_j as the orthogonal complement of V_j in V_{j+1}. It follows that

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j.$$

These spaces W_j will satisfy the same dilation property as the V_j's, i.e.,

$$\psi(x) \in W_j \Longleftrightarrow \psi(2x) \in W_{j+1}.$$

To obtain an orthonormal basis $\{\psi_{j,\,k}(x)\}_{j,\,k\in\mathbb{Z}}$ for $L^2(\mathbb{R})$, we can use the fact that, via the Fourier transform,

$$\hat{\phi}(2\omega) = H_0(\omega)\hat{\phi}(\omega), \text{ a.e. } \omega \in \hat{\mathbb{R}}^n, \tag{1.10}$$

where H_0 is a 1-periodic function [23]. A function ϕ that can be written as 1.10 is said to be **refinable**.

In this case, a choice of $\phi \in W_0$ that will generate a wavelet orthonormal basis is

$$\hat{\psi}(\omega) = \overline{H_0\left(\frac{\omega}{2}+\frac{1}{2}\right)}e^{-\pi i\omega}\hat{\phi}\left(\frac{\omega}{2}\right).$$

In this thesis, we are mainly concerned with conditions that guarantee the existence of a frame multiresolution analysis. In Chapter 2, we will obtain a condition similar to (1.10) for a frame generated by composite dilations and translations of a scaling function ϕ. In Chapter 3, we focus on obtaining systems that form a system close to a MRA for square integrable functions defined on spaces of homogeneous type including, but not limited to Euclidean spaces.

1.3 Thesis Contribution

The main goal of this thesis is to generalize the diffusion wavelets of Coifman and Maggioni. First, in Chapter 2, we obtain sufficient conditions for Composite Frames as part of the exploratory process to define Diffusion Wavelet Frames and

Composite Diffusion Frames. The idea of using a composition of dilation operators was inspired by shearlets, which were introduced by G. Kutyniok et al. [55], and have generated important developments in recent years [30, 43, 54, 67, 68]. Shearlets are a family of wavelets constructed via the composition of dyadic dilations and shear transformations, which pick up directionality in data sets. Shearlets are a particular case of Wavelets with Composite Dilations, a family of affine systems that can provide flexibility and efficiency for function representation [56]. That type of flexibility, and hence, adaptability to various data sets is exactly what we are seeking in the settings of graphs and manifolds by generalizing diffusion wavelets. The theory of this generalization is established in Chapter 3. We begin by constructing Diffusion Frames and Diffusion Wavelet Frames, which broaden the type of representation system (frame instead of ONB) and allow us to forego of the orthogonalization process that makes diffusion wavelets computationally expensive. Then, we introduce the theory of Composite Diffusion Frames, which are obtained by applying a composition of dilation operators to a diffusion frame. We argue that composite diffusion frames defined in this way could be used to define a notion of directionality on graphs and functions defined on graphs. At this stage, we present diffusion wavelet frames and composite diffusion frames only on a theoretical level and briefly discuss a strategy to obtain Diffusion Shearlets. Numerical examples are in progress, but we will not discuss them in this thesis.

In Chapter 4, our medical application is an effective illustration of the usefulness of harmonic analysis tools on domains other than subsets of $\mathbb{R}^n$. This application is also an example of the type of problems for which composite diffusion frames may

be useful in the future. We will come back to this remark in more detail later, in the conclusion of Chapter 4. We focus on the detection of reticular anomalies in autofluorescence retinal images taken by Denise Cunningham from the National Eye Institute, with the goal of diagnosing Age-related Macular Degeneration (AMD), the leading cause of vision loss in elderly patients in industrialized nation [63]. The novel method that we propose detects anomalies in images in two steps. First, we perform feature enhancement and dimensionality reduction using Laplacian Eigenmaps (LE) of Belkin and Niyogi [2]. Although LE has existed for a long time, using this algorithm in the context of detecting anomalies in the retina is a novel approach. LE belongs in the family of kernel based analysis techniques for manifold learning, which also include Kernel PCA [91], Locally Linear Embedding (LLE) [89], Diffusion Maps [26], Diffusion Wavelets [27, 64] etc. Many of these methods can be related to the family of diffusion methods that we construct in Chapter 3. Given the set of vectors $\mathcal{X} = \{x_1, \cdots x_N\}$, $x_i \in \mathbb{R}^D$, where D is large, all these methods start by 1) representing the data in form of a graph, where nodes represent the data vectors and edges represent some arbitrary relationship between pairs of vectors, 2) storing the graph in the adjacency matrix, 3) designing a kernel that captures some affinity or similarity between points of the graph. The methods can differ vastly on the kernel definition but most try to recover the underlying data manifold by representing the data in terms of the most significant eigenvectors of the kernel matrix or operator. In [59], the authors show that all kernel-based techniques are just special cases of kernel PCA. In Chapter 4, we also show how this applies to Laplacian Eigenmaps. Now, after performing the LE-step on retinal images, we obtain Laplacian eigenimages on which we perform

anomaly classification using Vectorized Matched Filtering algorithm (VMF). VMF is a novel approach, based on matched filtering, that views the retinal images as a data cubes and aggregates detections of anomalies on individual images into a single detection.

Chapter 2

Sufficient Conditions for Composite Frames

2.1 Introduction

In this chapter, we give sufficient conditions to obtain Composite Frames (or Frame $\mathbf{G}_A\mathbf{G}$-MRA). This is a frame multiresolution analysis generated by applying successive dilations based on elements of subgroups $\mathbf{G}_A$ and $\mathbf{G}$ of the general linear group to a function $\phi \in L^2(\mathbb{R}^n)$ satisfying certain requirements. As we stated in the introduction, our main goal in studying systems with composite dilations is to get an insight on how the same idea could be used to define more flexible, possibly directional, diffusion representation systems.

Our construction is related to Composite Wavelets (or Wavelets with Composite Dilations), which were introduced by Guo et al. in [56] as a class of affine systems of the form

$$\mathcal{A}_{\mathbf{G}_A\mathbf{G}} = \{D_A D_B T_k \psi : k \in \mathbb{Z}^n,\, B \in \mathbf{G},\, A \in \mathbf{G}_A\}\,, \tag{2.1}$$

where $\psi \in L^2(\mathbb{R}^n)$, T_k is translation by an integer k, D_A and D_B are dilation operators, and $\mathbf{G}_A$, $\mathbf{G}$ are countable subgroups of $GL_n(\mathbb{R})$, the general linear group of degree n over $\mathbb{R}$. These affine systems are a generalization of the traditional wavelet system $\left\{2^{-j/2}\psi\left(2^{-j/2}x - k\right) : x \in \mathbb{R},\, j,\, k \in \mathbb{Z}\right\}$.

Composite wavelets are an umbrella for numerous **directional** representation systems such as contourlets [37], curvelets [16], shearlets, [55, 73], which are extensions of traditional wavelets. These systems target applications in which they optimally, i.e., sparsely, represent functions with certain singularities or particular geometric features, based on the following criteria [37]:

- **Multiresolution.** Constituents of the representation approximate the data at successive, coarse to finer, resolutions.
- **Localization.** Constituents of the representation system are localized both in space and frequency domains.
- **Critical sampling.** Constituents of the representation must form a sparse system, i.e., a system with low redundancy for targeted applications.
- **Directionality.** Constituents of the representation must capture various orientations in the data.
- **Anisotropy.** Constituents of the representation must capture various elongated shapes with different aspect ratios.

The latter two properties are the key difference between directional representations and traditional wavelets, which contain isotropic elements occurring at all scales and locations. Directional representations are particularly useful for images. For example, contourlets are made of a discrete-domain multiresolution and multi-direction tight frame that efficiently approximates images made of smooth regions separated by smooth boundaries [37]. By finding a common framework for all these representations, not only does [56] provide a beautiful mathematical theory, but also, it allows

experimentation with specific parameters, which could seamlessly lead to representations with even more flexibility and ability to capture specific features of the data. The crux of [56] is to establish conditions on $\mathbf{G}_A$ and $\mathbf{G}$ such that the system $\mathcal{A}_{\mathbf{G}_A\mathbf{G}}$ is an orthonormal wavelet (or multiwavelet) basis or, more generally, a Parseval frame for $L^2(\mathbb{R}^n)$. In our construction, we take these conditions for granted. In particular, we use the idea of shift-invariant systems and the admissibility conditions on $\mathbf{G}_A$ and $\mathbf{G}$ in [56].

Our work differs from [56] mainly by its focus on constructing a frame multiresolution analysis of scaling functions instead of orthonormal wavelets. In particular, our main result, Theorem 2.14, gives conditions, most importantly, that of being **refinable** (2.5), that a function ϕ must satisfy to generate a frame $\mathbf{G}_A\mathbf{G}$-MRA. This theorem settles a case that was, so far, only stated as true without poof or proved for specific orthonormal bases [55, 56]. Theorem 2.14 is also attractive in that the proof uses technical arguments that give a better understanding of how the dilations interact to yield a frame $\mathbf{G}_A\mathbf{G}$-MRA.

This chapter is organized as follows. In Section 2.2, we begin by introducing the terminology used throughout the chapter. Then, we provide a few important results and summarize some important developments in directional methods. In Section 2.3, we present the admissibility conditions for sets of matrices used as dilation matrices in the frame $\mathbf{G}_A\mathbf{G}$-MRA and also give the formal definition of a frame $\mathbf{G}_A\mathbf{G}$-MRA. Then, to prove our main theorem, we assume that there exists a function ϕ whose integer translations and dilations by admissible matrices forms a semi-orthogonal Parseval frame for the space of square integrable functions on $\mathbb{R}^n$ and show that

whenever ϕ is refinable we obtain a frame $\mathbf{G}_A\mathbf{G}$-MRA. This condition of self-similarity for ϕ is related to others in classical wavelets theory [10, 11, 23, 79]. In Section 2.4, we produce an example of a frame based on the partial mollification of a 2-D Haar wavelet. Our frame differs from a $\mathbf{G}_A\mathbf{G}$-MRA only on a small set and have better smoothness properties. Finally, we summarize our results in the conclusion.

2.2 Background

2.2.1 Terminology

In this section, we formally introduce the operators used to build affine systems with composite dilations. Since we will use both their time and frequency properties, we provide useful commutative relations as well as formulas that describe how the Fourier transform acts on these operators.

Recall that we view $x \in \mathbb{R}^n$ as a column vector in space or time and $\omega \in \hat{\mathbb{R}}^n$ as a row vector in the frequency domain. We denote by $GL_n(\mathbb{R})$ the **linear group of degree** n, i.e., the set of $n \times n$ invertible matrices over the field $\mathbb{R}$ with the operation of ordinary matrix multiplication and by $SL_n(\mathbb{R})$ the **special linear group of degree** n, i.e., the set of $n \times n$ matrices in $GL_n(\mathbb{R})$ with determinant 1.

Definition 2.1. Let $f \in L^2(\mathbb{R}^n)$, $x \in \mathbb{R}^n$. We can define the following unitary operators on $L^2(\mathbb{R}^n)$:

- The **translation** of f by y, $T_y : L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n)$, where $y \in \mathbb{R}^n$, is given by:

$$(T_y f)(x) = f(x - y).$$

- The **dilation** of f by A, $D_A : L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n)$, where $A \in GL_n(\mathbb{R})$, is given by

$$(D_A f)(x) = |\det A|^{-1/2} f\left(A^{-1}x\right).$$

In particular, when $A = 2$, $(D_2 f)(x) = \frac{1}{\sqrt{2}} f\left(\frac{x}{2}\right)$ is the standard dyadic dilation, and **G** is the trivial group, we have the classical wavelet system:

$$\left\{2^{-j/2}\psi\left(2^{-j/2}x - k\right) : j,\, k \in \mathbb{Z}\right\}. \tag{2.2}$$

Next, we examine how translations and dilations commute with another. The following results can be found in standard books on wavelets on $\mathbb{R}^n$. We show the details of the proofs as it familiarizes us with common techniques when dealing with dilation and translation operators.

Proposition 2.2. *Let A, $B \in GL_n(\mathbb{R})$, $y \in \mathbb{R}^n$. We have*

(i) $D_A^{-1} = D_{A^{-1}}$.

(ii) $D_A D_B = D_{AB}$.

(iii) $D_A T_y = T_{A^{-1}y} D_A$.

Proof. The formulas follow from simple linear algebra manipulations. Let $f \in L^2(\mathbb{R}^n)$.

(i) For the first formula,

$$\begin{aligned} D_{A^{-1}}\left[D_A f\right](x) &= |\det(A)|^{1/2}\left[D_A f\right](Ax) \\ &= |\det(A)|^{1/2}\left(|\det(A)|^{-1/2} f(A^{-1}Ax)\right) \\ &= f(x). \end{aligned}$$

Hence $D_{A^{-1}} = D_A^{-1}$.

(ii) For the second formula,

$$\begin{aligned} D_A D_B\left[f\right](x) &= |\det(A)|^{-1/2}|\det(B)|^{-1/2} f(B^{-1}A^{-1}x) \\ &= |\det(AB)|^{-1/2} f((AB)^{-1}x) \\ &= D_{AB}\left[f\right](x). \end{aligned}$$

(iii) Last,

$$\begin{aligned} \left[D_A T_y f\right](x) &= D_A\left[f\right](x-y) \\ &= |\det A|^{-1/2} f(A^{-1}(x-y)) \\ &= |\det A|^{-1/2} T_{A^{-1}y} f(A^{-1}x) \\ &= \left[T_{A^{-1}y} D_A f\right](x). \end{aligned}$$

□

Let us introduce two unitary operators that will simplify certain frequency domain manipulations.

Definition 2.3. a) Let $y \in \mathbb{R}^n$, $A \in GL_n(\mathbb{R})$ and $g \in L^2(\hat{\mathbb{R}}^n)$. The **Fourier domain dilation** $\hat{D}_A$ of g is

$$\hat{D}_A\left(g\right)(\omega) = |\det A|^{1/2} g\left(\omega A\right),\ \omega \in \hat{\mathbb{R}}^n.$$

b) The operator $M_y : L^2(\hat{\mathbb{R}}^n) \longrightarrow L^2(\hat{\mathbb{R}}^n)$ is the **modulation** of g by $y \in \mathbb{R}^n$, given by:

$$\left(M_y g\right)(\omega) = e^{-2\pi i \omega y} g(\omega),\ \omega \in \hat{\mathbb{R}}^n.$$

Remark 2.4. Observe that the exponent in the modulation operator M_y contains a

negative sign. This definition is different from the usual definition of modulation, but it will yield more obvious commutation relationships between operators in the Fourier domain, and will spare us from the task of keeping track of a negative sign in future calculations.

We proceed to studying the action of Fourier transform on the operators defined above.

Proposition 2.5. *Let $A \in GL_n(\mathbb{R})$, $B \in SL_n(\mathbb{R})$ and y, $k \in \mathbb{R}^n$. Then,*

(i) $\mathscr{F} D_A = \hat{D}_A \mathscr{F}$.

(ii) $\mathscr{F} T_y = M_y \mathscr{F}$.

(iii) $\mathscr{F} D_A^j D_B = \hat{D}_{A^j} \hat{D}_B \mathscr{F}$.

Proof. For the first and second formulas, substitution in the Fourier integral is the main tool. Let $f \in L^2(\mathbb{R}^n)$.

(ii) We have

$$\begin{aligned}
\mathscr{F}[D_A f](\omega) &= \int_{\mathbb{R}^n} |\det A|^{-1/2} f(A^{-1}x) e^{-2\pi i \omega x} \mathrm{d}x \\
&= |\det A|^{-1/2} \int_{\mathbb{R}^n} f(u) e^{-2\pi i \omega A u} |\det A| \mathrm{d}u \\
&= |\det A|^{1/2} \hat{f}(\omega A) \\
&= \hat{D}_A \hat{f}(\omega).
\end{aligned}$$

(ii) Here,

$$\begin{aligned}\mathscr{F}\left[D_A f\right](\omega) &= \int_{\mathbb{R}^n} f(x-k)e^{-2\pi i\omega x}\mathrm{d}x \\ &= \int_{\mathbb{R}^n} f(u)e^{-2\pi i\omega(u+y)}\mathrm{d}u \\ &= e^{-2\pi i\omega y}\int_{\mathbb{R}^n} f(u)e^{-2\pi i\omega u}\mathrm{d}u \\ &= M_y\hat{f}(\omega).\end{aligned}$$

(iii) For these equalities use Proposition 2.2 (ii) and Proposition 2.5 (i) as follows:

$$\begin{aligned}\mathscr{F}\left[D_A^j D_B f\right](\omega) &= \mathscr{F}\left[D_{A^j B} f\right](\omega) \\ &= \hat{D}_{A^j B}\mathscr{F}\left[f\right](\omega) \\ &= \hat{D}_{A^j}\hat{D}_B\mathscr{F}\left[f\right](\omega),\end{aligned}$$

where the last equality is obtained in exactly the same way as we obtained 2.2 (ii). □

2.2.2 Shift-Invariant Spaces

Next, we examine shift-invariant spaces. Before we describe the importance of these spaces in our work, we state the definition.

Definition 2.6. A $\mathbb{Z}^n$-**invariant space** (or **shift-invariant space**) of $L^2(\mathbb{R}^n)$ is a closed subspace $V \subset L^2(\mathbb{R}^n)$ such that $T_k V = V$ for each $k \in \mathbb{Z}^n$. Let $\phi \in L^2(\mathbb{R}^n) \setminus \{0\}$. We denote by $\langle\phi\rangle$ the invariant shift space generated by ϕ:

$$\langle\phi\rangle = \overline{\text{span}}\left\{T_k\phi : k \in \mathbb{Z}^n\right\}.$$

We get an idea of the importance of shift-invariant spaces by observing Equation (2.2) and noting that, for traditional wavelets, we work on the shift invariant space

generated by a function $\psi \in L^2(\mathbb{R}^n)$ under the shifts of $\mathbb{Z}^n \subset \mathbb{R}^n$. Observing that $\mathbb{Z}^n$ is actually the semi-direct product of the trivial group in $SL_n(\mathbb{R})$ and $\mathbb{Z}^n$ is the bridge between shift-invariant spaces for classical wavelet theory and shift-invariant spaces for composite wavelets. For composite wavelets, we consider $L^2(\mathbb{R}^n)$-functions with shifts by the semi-direct product of $\mathbb{Z}^n$ and a discrete, and often non-abelian group, $\mathbf{G}$. In [79], [10], and [11], the authors study wavelets with composite dilations, where $\mathbf{G}$ is a crystallographic group, i.e., a group of isometries such as rotations or reflections on $\mathbb{R}^n$. The example that we give in Section 2.4 will also feature such a group, which naturally preserves the MRA structure of the Haar wavelet.

Remark 2.7. From now on, we only consider finite groups $\mathbf{G}$, but note that one can certainly define shift-invariant spaces for any countable group $\mathbf{G}$, as is done when defining shearlets.

Definition 2.8. Let $\mathbf{G}$ be a finite subgroup of $SL_n(\mathbb{Z})$ and let $M \ltimes N$ denote the semi-direct product of two groups M and N. The $\mathbf{G} \ltimes \mathbb{Z}^n$-**invariant spaces** are the closed subspaces $V \subset L^2(\mathbb{R}^n)$ for which $D_B T_k V = V$ for any pair $(B, k) \in \mathbf{G} \ltimes \mathbb{Z}^n$. Let $\phi \in L^2(\mathbb{R}^n)$. The $\mathbf{G} \ltimes \mathbb{Z}^n$-invariant spaces generated by ϕ, denoted $\langle\langle \phi \rangle\rangle$, are defined as

$$\langle\langle \phi \rangle\rangle = \overline{\text{span}} \{ D_B T_k \phi : B \in \mathbf{G},\, k \in \mathbb{Z}^n \}.$$

Definition 2.9. Let $\phi \in L^2(\mathbb{R}^n)$ and let $\mathbf{G}$ be a finite subgroup of $SL_n(\mathbb{Z})$. The set $\Phi_{\mathbf{G}} = \{ D_B T_k \phi : B \in \mathbf{G}, k \in \mathbb{Z}^n \}$ is a **semi-orthogonal Parseval frame** for the

$\mathbf{G} \ltimes \mathbb{Z}^n$-invariant subspace $\langle\langle\phi\rangle\rangle$ if $\{T_k\phi : k \in \mathbb{Z}^n\}$ is a Parseval frame for $\langle\phi\rangle$ and

$$\langle\langle\phi\rangle\rangle = \bigoplus_{B\in\mathbf{G}} D_B \langle\phi\rangle ,$$

i.e., $D_B T_k\phi \perp D_{B'} T_{k'}\phi$ for any B, $B' \in \mathbf{G}$, $B \neq B'$ and k, $k' \in \mathbb{Z}^n$. We also say that ϕ generates the semi-orthogonal Parseval frame $\Phi_{\mathbf{G}}$ in this case.

Note that Definition 2.9 can be extended to any frame and the condition of semi-orthogonality can be removed for more general shift-invariant spaces. However, as we will show in our example, adding those conditions simplify the construction of a frame $\mathbf{G}_A\mathbf{G}$-MRA.

2.2.3 $\mathbf{G}_A\mathbf{G}$-Multiresolution Analysis

Before we define the $\mathbf{G}_A\mathbf{G}$-MRA, we briefly discuss conditions on $\mathbf{G}_A$ and $\mathbf{G}$ that yield such a system.

Definition 2.10. A matrix A is an **expanding** matrix if all the eigenvalues λ of A satisfy the condition $|\lambda| > 1$.

Definition 2.11. Let $\mathbf{G}$ be a finite subgroup of $SL_n(\mathbb{Z})$ and $\mathbf{G}_A = \{A^j : j \in \mathbb{Z}\} \subset GL_n(\mathbb{R})$. We say that $A \in \mathbf{G}_A$ **normalizes** $\mathbf{G}$ if, for each $B \in \mathbf{G}$, $ABA^{-1} \in \mathbf{G}$.

Imposing the condition that A be an expanding matrix is not necessary to obtain an $\mathbf{G}_A\mathbf{G}$-MRA. However, when combined with the condition that A normalizes $\mathbf{G}$, it is a simple way to ensure that $\mathbf{G}_A\mathbf{G}$ meets the "admissibility condition" that guarantees the existence of a Parseval frame for $L^2(\mathbb{R}^n)$ of the form (2.1). For a more comprehensive understanding of admissibility conditions, refer to [56]. In our

case, this admissibility condition basically ensures that we obtain shift-invariance for fundamental domains of $\mathbb{R}^n$ and a nested sequence of MRA spaces.

Now, we are ready to define a $\mathbf{G}_A\mathbf{G}$-MRA.

Definition 2.12. Let $\mathbf{G}$ be a finite subgroup of $SL_n(\mathbb{Z})$ and $\mathbf{G}_A = \{A^j : j \in \mathbb{Z}\}$, where $A \in GL_n(\mathbb{Z})$ is an expanding matrix. Moreover, assume that A normalizes $\mathbf{G}$. We say that the sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R}^n)$ is a **$\mathbf{G}_A\mathbf{G}$-multiresolution analysis ($\mathbf{G}_A\mathbf{G}$-MRA)** if the following properties hold:

(i) $D_B T_k V_0 = V_0$, for all $B \in \mathbf{G}$, $k \in \mathbb{Z}^n$.

(ii) $V_j \subset V_{j+1}$ for each $j \in \mathbb{Z}$, where $V_j = D_A^{-j} V_0$.

(iii) $\bigcap\limits_{j\in\mathbb{Z}} V_j = 0$ and $\overline{\bigcup\limits_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^n)$.

(iv) There exists $\phi \in L^2(\mathbb{R}^n)$ such that $\Phi_{\mathbf{G}} = \{D_B T_k \phi : B \in \mathbf{G}, k \in \mathbb{Z}^n\}$ is a semi-orthogonal Parseval frame for V_0.

2.3 Sufficient Conditions

We establish sufficient conditions for a $\mathbf{G}_A\mathbf{G}$-MRA by the way of Lemma 2.13, which gives conditions that guarantee that $V_j \subset V_{j+1}$ for each $j \in \mathbb{Z}$. We make use of Proposition 1.10 and Proposition 1.7 in this first lemma.

Lemma 2.13. *Assume $\mathbf{G}$ is a finite subgroup of $SL_n(\mathbb{Z})$ and $\mathbf{G}_A = \{A^j : j \in \mathbb{Z}\}$, where $A \in GL_n(\mathbb{Z})$ is an expanding matrix. For $j \in \mathbb{Z}$, let V_j be defined as*

$$V_j = D_A^{-j} \overline{\operatorname{span}} \{\Phi_{\mathbf{G}}\} = \overline{\operatorname{span}} \left\{D_A^{-j} \Phi_{\mathbf{G}}\right\}. \tag{2.3}$$

Moreover, let $\phi \in L^2(\mathbb{R}^n)$ and assume that $\Phi_{\mathbf{G}} = \{D_B T_k \phi : B \in \mathbf{G}, k \in \mathbb{Z}^n\}$ is a

semi-orthogonal Parseval frame for V_0. Then, for $j \in \mathbb{Z}$,

(i) $D_A^{-j}\Phi_{\mathbf{G}}$ is a semi-orthogonal Parseval frame for V_j.

(ii) A function $f \in L^2(\mathbb{R}^n)$ belongs to V_j if and only if

$$f = \sum_{k\in\mathbb{Z}^n} \sum_{B\in\mathbf{G}} c_{k,B} D_{A^{-j}} D_B T_k \phi,$$

for some $\{c_{k,B}\}_{k\in\mathbb{Z}^n,\, B\in\mathbf{G}} \subset \ell^2(\mathbb{Z}^n)$.

(iii) A function $f \in L^2(\mathbb{R}^n)$ belongs to V_j if and only if there exist $L^2(\mathbb{T}^n)$ periodic functions $F_{B,\,j} \in L^2(\mathbb{T}^n)$, $B \in \mathbf{G}$, such that

$$\hat{f}\left(\omega A^j\right) = \sum_{B\in\mathbf{G}} F_{B,\,j}(\omega B)\, \hat{\phi}(\omega B). \tag{2.4}$$

Proof. (i) Since $\{T_k\phi : k \in \mathbb{Z}^n\}$ is a Parseval frame and D_A^{-j} is a unitary operator, by Proposition 1.10, we have that $\left\{D_A^{-j} T_k \phi : k \in \mathbb{Z}^n\right\}$ is a Parseval frame. For semi-orthogonality, we use again the fact that D_A is unitary, to obtain

$$\begin{aligned}
\left\langle D_A^{-j} D_B T_k \phi,\, D_A^{-j} D_{B'} T_{k'} \phi \right\rangle &= \left\langle (D_A^{-j})^* D_A^{-j} D_B T_k \phi,\, D_{B'} T_{k'} \phi \right\rangle \\
&= \left\langle D_B T_k \phi,\, D_{B'} T_{k'} \phi \right\rangle \\
&= 0,
\end{aligned}$$

whenever $B \neq B'$, by assumption.

(ii) This is a consequence of (i) combined with Proposition 1.7.

(iii) Using (ii), for each $f \in V_j$ we have

$$\begin{aligned}
\hat{f}(\omega) &= \mathscr{F}\left[\sum_{k\in\mathbb{Z}^n} \sum_{B\in\mathbf{G}} c_{k,B} D_A^{-j} D_B T_k \phi\right](\omega) \\
&= \sum_{k\in\mathbb{Z}^n} \sum_{B\in\mathbf{G}} c_{k,B} \mathscr{F}\left[D_A^{-j} D_B T_k \phi\right](\omega).
\end{aligned}$$

Making use of Proposition 2.5,

$$\begin{aligned}\hat{f}(\omega) &= \sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}} c_{k,B}\left[\hat{D}_A^{-j}\hat{D}_B M_k\hat{\phi}\right](\omega)\\ &= \hat{D}_A^{-j}\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}} c_{k,B}\left[\hat{D}_B M_k\hat{\phi}\right](\omega).\end{aligned}$$

Then,

$$\begin{aligned}\hat{f}\left(\omega A^j\right) &= |\det A|^{-j/2}\hat{D}_A^j\hat{f}(\omega)\\ &= |\det A|^{-j/2}\hat{D}_A^j\left(\hat{D}_A^{-j}\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}} c_{k,B}\left[\hat{D}_B M_k\hat{\phi}\right](\omega)\right)\\ &= |\det A|^{-j/2}\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}} c_{k,B}\left[\hat{D}_B M_k\hat{\phi}\right](\omega)\\ &= |\det A|^{-j/2}\sum_{B\in\mathbf{G}}\left(\sum_{k\in\mathbb{Z}^n} c_{k,B}e^{-2\pi i\omega Bk}\right)\hat{\phi}(\omega B).\end{aligned}$$

Now, for each $j\in\mathbb{Z}$, $B\in\mathbf{G}$, define

$$F_{B,j}(\omega) = |\det A|^{-j/2}\sum_{k\in\mathbb{Z}^n} c_{k,B}e^{-2\pi i\omega Bk}.$$

To verify that $F_{B,\,j}$ is in $L^2(\mathbb{T}^n)$, we compute

$$\begin{aligned}\int_{\mathbb{T}^n}|F_{B,j}(\omega)|^2\mathrm{d}\omega &= \int_{\mathbb{T}^n}\left||\det A|^{-j/2}\sum_{k\in\mathbb{Z}^n} c_{k,B}e^{-2\pi i\omega Bk}\right|^2\mathrm{d}\omega\\ &= |\det A|^{-j}\int_{\mathbb{T}^n}\sum_{k\in\mathbb{Z}^n}\sum_{k'\in\mathbb{Z}^n} c_{k,B}\overline{c_{k',\,B}}e^{-2\pi i\omega Bk}e^{2\pi i\omega Bk'}\mathrm{d}\omega.\end{aligned}$$

Now $\{c_{k,\,B}\}\subset\ell^2(\mathbb{Z}^n)$, so

$$\begin{aligned}\int_{\mathbb{R}^n}|F_{B,j}(\omega)|^2\mathrm{d}\omega &= |\det A|^{-j}\sum_{k\in\mathbb{Z}^n}\sum_{k'\in\mathbb{Z}^n} c_{k,B}\overline{c_{k',\,B}}\int_{\mathbb{T}^n}e^{-2\pi i\omega Bk}e^{2\pi i\omega Bk'}\mathrm{d}\omega\\ &= |\det A|^{-j}\sum_{k\in\mathbb{Z}^n}|c_{k,B}|^2\int_{\mathbb{T}^n}\left|e^{-2\pi i\omega Bk}\right|^2(\omega)\,\mathrm{d}\omega,\end{aligned}$$

since, for each B, $\left\{e^{-2\pi i\omega Bk}\right\}_{k\in\mathbb{Z}}$ form an orthonormal basis for $L^2(\mathbb{T}^n)$. Thus

$$\int_{\mathbb{R}^n} |F_{B,j}(\omega)|^2 \mathrm{d}\omega = |\det A|^{-j} \sum_{k\in\mathbb{Z}^n} |c_{k,B}|^2 < \infty.$$

Hence, we have that (2.4) holds with $F_{B,\,j}(\omega) = |\det A|^{-j/2} \sum_{k\in\mathbb{Z}^n} c_{k,B} e^{-2\pi i\omega Bk}$.

For the other direction, suppose that $f \in L^2(\mathbb{R}^n)$ and there exists a function $F_{B,\,j}$ such that Equation (2.4) is satisfied. Let $d_{B,k}$ be the Fourier coefficients of $F_{B,j}$. If we define $c_{B,k} = |\det A|^{j/2} d_{B,k}$, then we get $f = \sum_{k\in\mathbb{Z}^n} \sum_{B\in\mathbf{G}} c_{k,B} D_{A^{-j}} D_B T_k \phi$. Hence, by (ii), $f \in V_j$.

□

2.3.1 Main Result

We obtain sufficient conditions for composite frames or frame $\mathbf{G}_A\mathbf{G}$-MRA by proving that, given our set of assumptions, each property in Definition 2.12 holds. Our argument uses a few classical tools from Harmonic Analysis ([23], [33]) as well new approaches to deal with a few technicalities arising from our more complex setting.

Theorem 2.14. *(Sufficient Conditions for frame $\mathbf{G}_A\mathbf{G}$-MRA.) Let $\mathbf{G}$ be a finite subgroup of $SL_n(\mathbb{Z})$ and $\mathbf{G}_A = \{A^j : j \in \mathbb{Z}\}$, where $A \in GL_n(\mathbb{Z})$ is an expanding matrix. Assume that A* normalizes *$\mathbf{G}$. Assume also that $\phi \in L^2(\mathbb{R}^n)$ and that $\Phi_{\mathbf{G}} = \{D_B T_k \phi : B \in \mathbf{G}, k \in \mathbb{Z}^n\}$ is a semi-orthogonal Parseval frame for V_0. Define*

$$V_j = D_A^{-j} V_0.$$

If $\hat{\phi}$ is continuous, and uniformly bounded on a neighborhood of zero, $\hat{\phi} \neq 0$, and there

exist $\mathbb{T}^n$-periodic functions $H_B \in L^\infty(\mathbb{T}^n)$ such that

$$\hat{\phi}(\omega) = \sum_{B \in \mathbf{G}} H_B\left(\omega B A^{-1}\right) \hat{\phi}\left(\omega B A^{-1}\right), \tag{2.5}$$

then ϕ generates a semi-orthogonal Parseval frame $\mathbf{G}_A\mathbf{G}$-multiresolution analysis.

Proof. (i) To show that our system satisfies (i) in Definition 2.12, we simply observe the fact that $\Phi_{\mathbf{G}}$ is a semi-orthogonal Parseval frame for V_0 implies

$$V_0 = \overline{\text{span}}\left\{\Phi_{\mathbf{G}}\right\},$$

which is, by assumption, an $\mathbf{G} \ltimes \mathbb{Z}^n$-invariant subspace of $L^2(\mathbb{R}^n)$, i.e., $D_B T_k V_0 = V_0$.

(ii) Let $f \in V_j$, $j \in \mathbb{Z}$. We want to show that this implies $f \in V_{j+1}$, or equivalently, via Lemma 2.13, that there exist $\mathbb{T}^n$-periodic functions $F_{B,\,j+1} \in L^2(\mathbb{T}^n)$ such that Equation (2.4) holds.

Employing (iii) of Lemma 2.13 yields

$$\begin{aligned} \hat{f}(\omega A^{j+1}) &= \hat{f}((\omega A)A^j) \\ &= \sum_{B' \in \mathbf{G}} F_{B',j}(\omega A B')\, \hat{\phi}(\omega A B'), \end{aligned}$$

where $F_{B',j}(\omega)$ is a $\mathbb{T}^n$-periodic, $L^2(\mathbb{T}^n)$ function.

Using the assumption (2.5),

$$\begin{aligned} \hat{f}\left(\omega A^{j+1}\right) &= \sum_{B' \in \mathbf{G}} F_{B',j}(\omega A B') \left(\sum_{B'' \in \mathbf{G}} H_{B''}\left(\omega A B' B'' A^{-1}\right) \hat{\phi}\left(\omega A B' B'' A^{-1}\right) \right) \\ &= \sum_{B' \in \mathbf{G}} F_{B',j}(\omega A B') \left(\sum_{B'' \in \mathbf{G}} H_{B''}\left(\omega A B' B'' A^{-1}\right) \hat{\phi}\left(\omega A B' B'' A^{-1}\right) \right) \end{aligned} \tag{2.6}$$

Now, since $\mathbf{G}_A$ normalizes $\mathbf{G}$, we have that $A'B'B''A^{-1} \in \mathbf{G}$. Let $B = AB'B''A^{-1}$. We claim that for distinct B'_1, B'_2 in the first sum, $A'B'_1B''A^{-1}$ and $A'B'_2B''A^{-1}$

are distinct. We show this by contradiction. Suppose that the claim is false, i.e., $A'B'_1B''A^{-1} = A'B'_2B''A^{-1}$. Then, we would have

$$\begin{aligned} A'B'_1B''A^{-1} &= A'B'_2B''A^{-1} \\ B'_1B'' &= B'_2B'' \\ B'_1 &= B'_2, \end{aligned}$$

a contradiction.

Therefore, we have

$$\begin{aligned} \hat{f}\left(\omega A^{j+1}\right) &= \sum_{B'\in\mathbf{G}} F_{B',j}\left(\omega AB'\right)\left(\sum_{B\in\mathbf{G}} H_{B''}\left(\omega B\right)\hat{\phi}\left(\omega B\right)\right) \\ &= \sum_{B\in\mathbf{G}}\left(\sum_{B'\in\mathbf{G}} F_{B',j}\left(\omega AB'\right)H_{B''}\left(\omega B\right)\right)\hat{\phi}\left(\omega B\right), \end{aligned}$$

where $B'' = A^{-1}(B')^{-1}BA$.

Let

$$F_{B,\,j+1}(\omega) = \sum_{B'\in\mathbf{G}} F_{B',j}\left(\omega AB'B^{-1}\right)H_{B''}\left(\omega\right).$$

With $F_{B',j}$ and $H_{B''}$ $\mathbb{T}^n$-periodic, $F_{B,\,j+1}$ is $\mathbb{T}^n$-periodic. We obtain $F_{B,\,j+1} \in L^2\left(\mathbb{R}^n\right)$ via the following calculation

$$\begin{aligned} \int_{\mathbb{T}^n} |F_{B,\,j+1}(\omega)|^2 \mathrm{d}\omega &= \int_{\mathbb{T}^n}\left|\sum_{B'\in\mathbf{G}} F_{B',j}\left(\omega AB'B^{-1}\right)H_{B''}\left(\omega\right)\right|^2 \mathrm{d}\omega \\ &\leq \|H_{B''}\|^2_{L^\infty(\mathbb{T}^n)}\int_{\mathbb{T}^n}\left|\sum_{B'\in\mathbf{G}} F_{B',j}\left(\omega AB'B^{-1}\right)\right|^2 \mathrm{d}\omega \\ &\leq \|H_{B''}\|^2_{L^\infty(\mathbb{T}^n)}\#|\mathbf{G}|\int_{\mathbb{T}^n}\sum_{B'\in\mathbf{G}}\left|F_{B',j}\left(\omega AB'B^{-1}\right)\right|^2 \mathrm{d}\omega, \end{aligned}$$

by Minkowski's inequality. Consequently,

$$\int_{\mathbb{T}^n} |F_{B,\,j+1}(\omega)|^2 \mathrm{d}\omega \leq \|H_{B''}\|^2_{L^\infty(\mathbb{T}^n)} \#|\mathbf{G}| \sum_{B'\in\mathbf{G}} \int_{\mathbb{T}^n} \left|F_{B',j}\left(\omega A B' B^{-1}\right)\right|^2 \mathrm{d}\omega < \infty.$$

The next step is to show that the intersection V_j, $j \in \mathbb{Z}$, is empty of that the closure of their union is the entire space $L^2(\mathbb{R}^n)$.

(iii) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$. Let $f \in \bigcap_{j\in\mathbb{Z}}$ and let $\varepsilon > 0$ be fixed. Recall that the set of compactly supported and continuous functions on $\mathbb{R}^n$, $\mathcal{S}(\mathbb{R}^n)$, is dense is $L^2(\mathbb{R}^n)$. Thus, there exists a function $f_c \in \mathcal{S}(\mathbb{R}^n)$, with compact support, $\mathrm{supp}\,(f_c) \subset B_R(0)$ for some $R > 0$, such that

$$\|f - f_c\|_{L^2(\mathbb{R}^n)} \leq \varepsilon/2.$$

Let P_j be the orthogonal projection operator onto V_j, $j \in \mathbb{Z}$. Then,

$$\|f - P_j f_c\|_{L^2(\mathbb{R}^n)} \;=\; \|P_j(f - f_c)\|_{L^2(\mathbb{R}^n)} \leq \|f - f_c\|_{L^2(\mathbb{R}^n)} \leq \varepsilon/2,$$

and therefore,

$$\|f\|_{L^2(\mathbb{R}^n)} \leq \varepsilon/2 + \|P_j f_c\|_{L^2(\mathbb{R}^n)}, \text{ for all } j \in \mathbb{Z}. \tag{2.7}$$

Let $\phi^B_{j,k} := D_A^{-j} D_B \phi$, $A \in \mathbf{G}$, $B \in \mathbf{G}$, $k \in \mathbb{Z}^n$, $j \in \mathbb{Z}$. Since, by assumption, $\Phi_{\mathbf{G}} = \left\{\phi^B_{0,k} : B \in \mathbf{G},\, k \in \mathbb{Z}^n\right\}$ is a Parseval frame for V_0, and consequently, by (i), $D_A^{-j}\Phi_{\mathbf{G}} = \left\{\phi^B_{j,k} : B \in \mathbf{G},\, k \in \mathbb{Z}^n\right\}$ is a Parseval frame for V_j, we have

$$\|P_j f_c\|^2_{L^2(\mathbb{R}^n)} = \sum_{k\in\mathbb{Z}^n} \sum_{B\in\mathbf{G}} \left|\left\langle f_c,\, \phi^B_{j,k}\right\rangle\right|^2.$$

Now,

$$\begin{aligned}
\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\left|\left\langle f_c,\,\phi^B_{j,\,k}\right\rangle\right|^2 &= \sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\left(\int_{B_R(0)} f_c(x)\phi^B_{j,\,k}(x)\mathrm{d}x\right)^2 \\
&\leq \sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\left(\int_{B_R(0)} |f_c(x)|\left|\phi^B_{j,\,k}(x)\right|\mathrm{d}x\right)^2 \\
&\leq \|f_c\|^2_{L^\infty(\mathbb{R}^n)}\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\left(\int_{B_R(0)}\left|\phi^B_{j,\,k}(x)\right|\mathrm{d}x\right)^2 \qquad (2.8)\\
&\leq \|f_c\|^2_{L^\infty(\mathbb{R}^n)}|B_R\,(0)|\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\int_{B_R(0)}\left|\phi^B_{j,\,k}(x)\right|^2\mathrm{d}x, \qquad (2.9)
\end{aligned}$$

where, for (2.9), we used the fact that $f_c \in \mathcal{S}$, and hence, has finite $L^\infty(\mathbb{R}^n)$ norm, and for (2.9), we used Cauchy-Schwartz.

Then,

$$\begin{aligned}
\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\left|\left\langle f_c,\,\phi^B_{j,\,k}\right\rangle\right|^2 &\leq \|f_c\|^2_{L^\infty(\mathbb{R}^n)}|B_R\,(0)||\det A|^{-j}\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\int_{B_R(0)}\left|\phi\left(B^{-1}A^{-j}x-k\right)\right|^2\mathrm{d}x \\
&= \|f_c\|^2_{L^\infty(\mathbb{R}^n)}|B_R\,(0)||\det A|^{-j}\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\int_{B_{\|A\|^{-j}R}(k)}|\det A|^j|\phi(y)|^2\mathrm{d}y \\
&= C_{f,\,R}\int_{S_{R,\,j}}|\phi(y)|^2\mathrm{d}y,
\end{aligned}$$

where $C_{f,\,\mathbf{G}} = \|f_c\|^2_{L^\infty(\mathbb{R}^n)}|B_R\,(0)||\mathbf{G}|$ and $S_{R,\,j} = \bigcup\sum_k B_{\|A\|^{-j}R}\,(k)$. Note that $C_{f,\,\mathbf{G}} < \infty$ since the support of f_c is of bounded measure and $\mathbf{G}$ is finite by assumption.

We want to show that if we can pick j large enough, we will obtain

$$\int_{S_{R,j}}|\phi(y)|^2\mathrm{d}y \leq \frac{\varepsilon^2}{4C_{f,R}}, \qquad (2.10)$$

and, therefore,

$$\sum_{k\in\mathbb{Z}^n}\sum_{B\in\mathbf{G}}\left|\left\langle f_c,\,\phi^B_{j,\,k}\right\rangle\right|^2 \leq \frac{\varepsilon^2}{4}. \qquad (2.11)$$

Write

$$\int_{S_{R,j}} |\phi(y)|^2 \mathrm{d}y = \int_{S_{R,j}\cap E} |\phi(y)|^2 \mathrm{d}y + \int_{S_{R,j}\cap E^c} |\phi(y)|^2 \mathrm{d}y, \tag{2.12}$$

where $E = B_t(0)$, $t > 0$. Then

$$\begin{aligned} \int_{S_{R,j}\cap E^c} |\phi(y)|^2 \mathrm{d}y &\leq \int_{E^c} |\phi(y)|^2 \mathrm{d}y \\ &= \int_{|y|\geq t} |\phi(y)|^2 \mathrm{d}y. \end{aligned}$$

Now, since ϕ is $L^2(\mathbb{R}^n)$, then ϕ^2 is $L^1(\mathbb{R}^n)$, and by the Weirstrass theorem, there exists $t_0 \in \mathbb{R}$ such that

$$\int_{|y|\geq t_0} |\phi(y)|^2 \mathrm{d}y \leq \frac{\varepsilon^2}{8C_{f,R}}. \tag{2.13}$$

Let m be the Lebesgue measure on $\mathbb{R}^n$. Fix $t = t_0$. Since $t < \infty$, observe that $E = B_t(0)$ is a set of finite measure and therefore E contains finitely many integers k. That is, for each $j \in \mathbb{Z}$, there exists a finite sequence of $\mathrm{K} \subset \mathbb{Z}$ such that

$$S_{R,j} \cap E = \left[\bigcup_{k\in \mathrm{K}} B_{\|A\|^{-j}R}(k)\right] \cap E.$$

Since A is expanding, $\|A\| > 1$, and it follows that $\|A\|^{-(j+1)} \leq \|A\|^{-j}$. Thus, for R fixed, $B_{\|A\|^{-(j+1)}R}(k) \subseteq B_{\|A\|^{-j}R}(k)$ and hence,

$$\left[\bigcup_{k\in \mathrm{K}} B_{\|A\|^{-(j+1)}R}(k) \cap E\right] \subseteq \left[\bigcup_{k\in \mathrm{K}} B_{\|A\|^{-j}R}(k) \cap E\right].$$

Consequently,

$$(S_{R,j+1} \cap E) \subseteq (S_{R,j} \cap E)$$

By a standard result of real analysis [90], since $S_{R,j} \cap E$, $j \in \mathbb{Z}$ is a sequence of

decreasing measurable sets, we have

$$\mathrm{m}\left(\bigcap_{j=1}^{\infty}(S_{R,j}\cap E)\right)=\lim_{j\longrightarrow\infty}\mathrm{m}\left(S_{R,j}\cap E\right). \tag{2.14}$$

Now, as $j\longrightarrow\infty$, $\mathrm{m}(B_{\|A\|^{-j}R}(k))\longrightarrow 0$, because $\|A\|>1$. Therefore, as $j\longrightarrow\infty$,

$$\mathrm{m}\left(\bigcup_{k\in\mathrm{K}}B_{\|A\|^{-j}R}(k)\cap E\right)\longrightarrow 0.$$

Applying (2.14), we obtain

$$\mathrm{m}\left(\bigcap_{j=1}^{\infty}(S_{R,j}\cap E)\right)=0.$$

In particular, as $j\longrightarrow\infty$, $\bigcap_{j=1}^{\infty}(S_{R,j}\cap E)$ contains only the integers $k\in\mathrm{K}$. Let $\chi_{S_{R,\,i}\cap E}(x)=1$ if $x\in S_{R,\,i}\cap E$ and $\chi_{S_{R,\,i}\cap E}(x)=0$ otherwise.

Using our recent calculations, we get

$$\int_{S_{R,\,J}\cap E}|\phi(y)|^2\mathrm{d}y \;=\; \int_{\mathbb{R}^n}|\phi(y)|^2\chi_{S_{R,\,i}\cap E}(y)\mathrm{d}y \tag{2.15}$$

Now, as $j\longrightarrow\infty$, if $y\notin\mathbb{Z}^n$, in particular, if y is not one of the $k\in\mathrm{K}$, $\chi_{S_{R,\,i}\cap E}\longrightarrow 0$. Therefore, using dominated convergence theorem, we have (2.15) goes to 0 as $j\longrightarrow\infty$. Hence, for j large enough, we have,

$$\int_{S_{R,\,J}\cap E}|\phi(y)|^2\mathrm{d}y \;=\; \frac{\varepsilon^2}{8C_{f,\,R}}. \tag{2.16}$$

Therefore, for j large enough, if we add (2.13) and (2.16), we obtain that (2.11) holds. Employing (2.11) in the right-hand side of (2.7) yields $\|f\|_{L^2(\mathbb{R}^n)}\leq\varepsilon$. Taking $\varepsilon\longrightarrow 0$, we finally obtain $f=0$.

(iv) Last, we show that $\overline{\bigcup_{j\in\mathbb{Z}}V_j}=L^2(\mathbb{R}^n)$. Let $\varepsilon>0$ and pick a function f such that $f\in\left(\bigcap_{j\in\mathbb{Z}}V_j\right)^{\perp}$. Using density of $L^2(\mathbb{R}^n)$ in $\mathcal{S}$ once more, we can find a function

$f_c \in \mathcal{S}$ such that $\|f - f_c\|_{L^2(\mathbb{R}^n)} \leq \varepsilon$. Then, on one hand, for $J = -j$, $j \in \mathbb{Z}$, we have

$$\|P_{-J} f_c\|_{L^2(\mathbb{R}^n)} \leq \varepsilon, \tag{2.17}$$

and one the other hand, the frame condition implies that

$$\|P_{-J} f_c\|^2_{L^2(\mathbb{R}^n)} \geq \sum_{B \in \mathbf{G}} \sum_{k \in \mathbb{Z}^n} \left| \left\langle f_c,\, \phi^B_{-J,k} \right\rangle \right|^2. \tag{2.18}$$

For the right hand side RHS of the last equation, we begin by computing some estimates.

First, using Plancherel formula,

$$\begin{aligned}
RHS &= \sum_{B \in \mathbf{G}} \sum_{k \in \mathbb{Z}^n} \left| \int_{\mathbb{R}^n} \hat{f}_c(\omega) \overline{\hat{D}_{A^{-J}} \hat{D}_B \hat{\phi}}\, \mathrm{d}\omega \right|^2 \\
&= \sum_{B \in \mathbf{G}} \sum_{k \in \mathbb{Z}^n} \left| \int_{\mathbb{R}^n} \hat{f}_c(\omega) |\det A|^{-J} \overline{\hat{\phi}(\omega B A^{-J})} e^{2\pi i \omega B A^{-J} k}\, \mathrm{d}\omega \right|^2 \\
&= \sum_{B \in \mathbf{G}} \sum_{k \in \mathbb{Z}^n} \left| \int_{\mathbb{T}^n} e^{2\pi i \omega B A^{-J} k} \sum_{l \in \mathbb{Z}^n} \hat{f}_c(\omega + B A^J l) \overline{\hat{\phi}(\omega B A^{-J} + l)}\, \mathrm{d}\omega \right|^2 \\
&= \sum_{B \in \mathbf{G}} \left| \int_{\mathbb{T}^n} \sum_{l \in \mathbb{Z}^n} \hat{f}_c(\omega + B A^J l) \overline{\hat{\phi}(\omega B A^{-J} + l)}\, \mathrm{d}\omega \right|^2 \\
&= \sum_{B \in \mathbf{G}} \sum_{k \in \mathbb{Z}^n} \left| \int_{\mathbb{R}^n} \overline{\hat{f}_c(\omega)} \hat{f}_c(\omega + B A^J k) \overline{\phi(\omega \hat{A}^{-J})} \hat{\phi}(\omega B A^{-J} + k)\, \mathrm{d}\omega \right|^2
\end{aligned} \tag{2.19}$$

Define

$$S_0 = \sum_{B \in \mathbf{G}} \left| \int_{\mathbb{R}^n} |\hat{f}_c(\omega)|^2 |\hat{\phi}(\omega B A^{-J})|^2\, \mathrm{d}\omega \right|^2$$

and

$$S_{Rest} = \sum_{B \in \mathbf{G}} \sum_{\substack{k \in \mathbb{Z}^n \\ k \neq 0}} \left| \int_{\mathbb{R}^n} \overline{\hat{f}_c(\omega)} \hat{f}_c(\omega + B A^J k) \overline{\phi(\omega \hat{A}^{-J})} \hat{\phi}(\omega B A^{-J} + k)\, \mathrm{d}\omega \right|^2.$$

Since $f_c \in \mathcal{S}$, there exists $C > 0$ such that

$$|\hat{f}_c(\omega)| \leq C \left(1 + |\omega|^2\right)^{-3/2}.$$

Then, we have

$$\begin{aligned} S_{Rest} &\leq \|\hat{\phi}\|^2_{L^\infty(\mathbb{R}^n)} \sum_{B \in \mathbf{G}} \sum_{\substack{k \in \mathbb{Z}^n \\ k \neq 0}} \left| \int_{\mathbb{R}^n} \overline{\hat{f}_c(\omega)} \hat{f}_c(\omega + BA^J l)\, \mathrm{d}\omega \right|^2 \\ &\leq C^2 \|\hat{\phi}\|^2_{L^\infty(\mathbb{R}^n)} \sum_{B \in \mathbf{G}} \sum_{\substack{k \in \mathbb{Z}^n \\ k \neq 0}} \left(1 + \left|\omega + A^J l\right|^2\right)^{-3/2} \left(1 + \left|\omega - A^J l\right|^2\right)^{-3/2} \mathrm{d}\omega. \end{aligned}$$

After applying after applying some upper bound estimates, we arrive at

$$S_{Rest} \leq C'' 2^{-J}, \tag{2.20}$$

where $C'' > 0$.

Combining (2.17), (2.18), (2.19) and (2.20), we obtain

$$S_0 \leq B\varepsilon^2 + C'' 2^{-J}. \tag{2.21}$$

Now, by assumption, $\hat{\phi}(\omega)$ is uniformly bounded and continuous at $\omega = 0$. Thus so, by the dominated convergence theorem, for $J \to \infty$, $S_0 \to |\hat{\phi}(0)|^2 \|f_c\|^2_{L^2(\mathbb{R}^n)}$. Therefore, for $C > 0$ independent of ε,

$$\|f_c\|_{L^2(\mathbb{R}^n)} \leq |\hat{\phi}(0)|^{-1} C\varepsilon. \tag{2.22}$$

Putting together $\|f - f_c\|_{L^2(\mathbb{R}^n)} \leq \varepsilon$ and (2.22), we arrive at

$$\|f\|_{L^2(\mathbb{R}^n)} \leq \varepsilon + \|f_c\|_{L^2(\mathbb{R}^n)} \leq \left(1 + C|\hat{\phi}(0)|^{-1}\right) \varepsilon,$$

where C is some finite positive constant. Letting $\varepsilon \to 0$, we obtain $f = 0$.

□

2.4 Example: Approximate $\mathbf{G}_A\mathbf{G}$-MRA

In this section, we present an example of composite dilation frame based on a Haar-like wavelet associated with the quincunx dilation in $\mathbb{R}^2$. This Haar wavelet is presented in [72] and also discussed briefly by Manning and Blanchard in the context of, respectively, composite MRA wavelets [79] and crystallographic Haar-type composite dilation wavelets [11], [10]. We chose to base our example on this Haar wavelet both because of its simplicity in illustrating a $\mathbf{G}_A\mathbf{G}$-MRA and also because it shows the type of difficulties that can arise when one tries to obtain more desirable properties for composite dilation systems. We begin by describing the Haar system and verify that this is exactly an example of $\mathbf{G}_A\mathbf{G}$-MRA. Then, we deform the scaling function ϕ of this Haar system to obtain a function ϕ^ε, $\varepsilon > 0$, smooth on one side of the boundary of the domain of ϕ. We show that ϕ^ε yields a frame for a pre-defined space V_0 and also generates a frame system which approximates a $\mathbf{G}_A\mathbf{G}$-MRA in a sense that we will define later.

Let A be the quincux matrix:

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \tag{2.23}$$

and $G = \{B_i : i = 0, \ldots, 7\}$ be the group of symmetries of the unit square, that is,

$$B_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad B_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

and for $i = 4, 5, 6, 7$, $B_i = -B_{i-4}$.

Now, consider the region $R_0 = \{(x_1, x_2) : 0 \le x_1 < 1/2, 0 \le x_2 \le x_1\}$, i.e., R_0 is the

triangular region with the vertices (0, 0), $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$. For $i = 1, \ldots, 7$, define $R_i = B_i R_0$. The resulting triangles are shown in Figure 2.1.

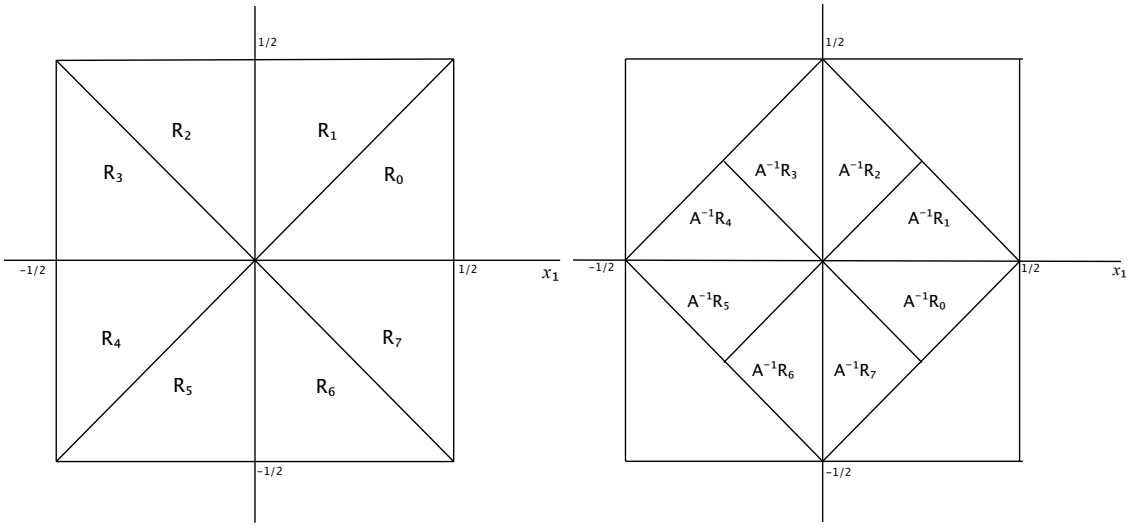


Figure 2.1: Fundamental domain R_0 of the scaling function ϕ and its images, on the left, under $B_1, \ldots, B_7$, and on the right, under $B_1, \ldots, B_7$, composed with A^{-1}.

Let ϕ be defined as

$$\phi = \sqrt{8}\chi_{R_0}. \tag{2.24}$$

We will use the following results, proved in [72].

Proposition 2.15. *Let ϕ be defined as in* (2.24). *Suppose A is the quincux matrix in* (2.23) *and $G = \{B_i : i = 1, \ldots, 7\}$ is the group of symmetries of the unit square.*

(i) The system $\Phi_{\mathbf{G}} = \{D_{B_i} T_k \phi : B_i \in \mathbf{G}, k \in \mathbb{Z}^2\}$ is an orthonormal basis for its closed linear span $V_0 \subset L^2(\mathbb{R}^2)$, which is comprised of all square integrable functions that are constant on each $\mathbb{Z}^2$-translate of the triangles R_i, $i = 0, \ldots, 7$.

(ii) Let $V_j = D_{A^{-j}} V_0$, $j \in \mathbb{Z}$, i.e., V_j contains all $L^2(\mathbb{R}^2)$-functions constant on

$A^{-j}\mathbb{Z}^2$ translates of the triangles $A^{-j}R_i$, $i = 0, \ldots, 7$. Then $V_j \subset V_{j+1}$, $j \in \mathbb{Z}$.

(iii) The spaces V_j, $j \in \mathbb{Z}$ form an $\mathbf{G}_A\mathbf{G}$-MRA with ϕ, which can be written as

$$\phi = \sqrt{2}\left(D_A^{-1}D_{B_1}\phi(x) + D_{A^{-1}}D_{B_6}T_{(1/2,\,1/2)}\phi(x)\right). \tag{2.25}$$

In Fourier domain, we have

$$\begin{aligned}\hat{\phi}(\omega) &= \sqrt{2}\left[\hat{D}_A^{-1}\hat{D}_{B_1}\hat{\phi}(\omega) + \hat{D}_{A^{-1}}\hat{D}_{B_6}M_{(1/2,\,1/2)}\hat{\phi}(\omega)\right] \\ &= \sqrt{2}\left[\hat{\phi}(\omega B_1 A^{-1}) + e^{-2\pi i\omega B_6 A^{-1}(1/2,\,1/2)}\hat{\phi}(\omega B_6 A^{-1})\right],\end{aligned}$$

i.e., ϕ meets condition (2.5) *with*

$$\begin{aligned}H_{B_1} &= \sqrt{2},\ \ H_{B_6} = \sqrt{2}e^{-2\pi i\omega B_6 A^{-1}(1/2,\,1/2)}, \\ H_{B_2} &= H_{B_3} = H_{B_4} = H_{B_5} = 0.\end{aligned}$$

Remark 2.16. We will not give a formal proof but just make a few observations. To verify (ii), first consider the images of the R_i in Figure 2.1, which show $V_0 \subset V_1$. From that observation and the definition of the V_j's, the conclusion that the V_j's form a nested sequence is obvious. Formula (2.25) can be verified by, again, looking at the right picture in Figure (2.1) and noting

$$R_0 = A^{-1}R_1 \cup \left[A^{-1}R_6 + \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}\right].$$

Next, we focus on showing that a "smooth" deformation ϕ^ε of ϕ on one side of the triangle R_0 generates a frame for V_0. For $x \in \mathbb{R}^n$, consider a "standard" mollifier

$$\eta(x) = \begin{cases} Ce^{\frac{1}{|x|^2-1}}, & |x| < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{2.26}$$

where C is chosen so that $\int_{\mathbb{R}^n} \eta(x)\,\mathrm{d}x = 1$.

Mollifiers, also known as approximations to identity or Friedrichs mollifiers after Kurt Otto Friedrichs who introduced them in [52], are smooth functions with many useful properties. In many cases, such as our present example, mollifiers are used to build sequences of smooth approximations to a given function via convolution. We will state properties pertinent to our proof later.

Given, $\varepsilon > 0$, let

$$\eta_\varepsilon(x) := \frac{1}{\varepsilon^2}\eta\left(\frac{x}{\varepsilon}\right).$$

It is easy to see that

$$\int_{\mathbb{R}^2} \eta_\varepsilon(x)\mathrm{d}x = 1 \text{ and } \operatorname{supp}(\eta_\varepsilon) \subset B_\varepsilon(0).$$

Now, we define the mollification f^ε of $f : U \longrightarrow \mathbb{R}$, where $U \subset \mathbb{R}^n$ and $f \in L^2(\mathbb{R}^n)$ locally, by the convolution

$$\begin{aligned} f^\varepsilon(x) &:= (\eta_\varepsilon * f)(x), \quad x \in U_\varepsilon \\ &= \int_{\mathbb{R}^2} \eta(y) f(x-y)\,\mathrm{d}y, \end{aligned}$$

where $U_\varepsilon := \{x \in U : \|x - \partial U\| > \varepsilon\}$.

Following these definitions, we can give a few useful properties of mollifiers. For the proofs, refer to [47].

Proposition 2.17. *Let f^ε be the mollification of $f : U \longrightarrow \mathbb{R}$, $U \subset \mathbb{R}^n$, $f \in L^1(\mathbb{R}^n)$ locally. Then*

(i) $f^\varepsilon \in C^\infty(U_\varepsilon)$,

(ii) $f^\varepsilon \longrightarrow f$ *a.e. as* $\varepsilon \longrightarrow 0$,

(iii) $\operatorname{supp}(f^\varepsilon) \subset \operatorname{supp}(f) + \overline{B_\varepsilon(0)} = \left\{x + y : x \in \operatorname{supp}(f), y \in \overline{B_\varepsilon(0)}\right\}$.

(iv) if $1 \leq p < \infty$ *and* $f \in L^p(U)$ *locally, then* $f^\varepsilon \longrightarrow f$ *in* $f \in L^p(U)$ *locally.*

Let $x = (x_1, x_2)$ and consider

$$\rho(x) = \begin{cases} \sqrt{8}, & 1/2 - \varepsilon/2 \leq x_1 \leq 1/2 + \varepsilon/2,\ -\varepsilon/2 \leq x_2 \leq x_1, \\ 0, & \text{otherwise.} \end{cases}$$

Define

$$\rho^\varepsilon(x) = \left(\eta_{\varepsilon/4} * \rho\right)(x),$$

and note that, by Proposition 2.17(i), (ii) and (iii) respectively,

1) $\rho^\varepsilon(x) \in C^\infty(\mathbb{R}^2)$.

2) $\operatorname{supp}(\rho^\varepsilon) = \{(x_1, x_2) : 1/2 - \varepsilon \leq x_1 \leq 1/2 + \varepsilon,\ -\varepsilon \leq x_2 \leq x_1\}$.

3) $\operatorname{supp}(\rho^\varepsilon) = \sqrt{8}$ on $\{(x_1, x_2) : -1/2 - \varepsilon/4 \leq x_1 \leq 1/2 + \varepsilon/4,\ -\varepsilon/4 \leq x_2 \leq x_1 + \varepsilon/4\}$.

This way of defining ρ^ε is just a particular case of cutoff or bump functions. Indeed, for given open sets Ω' compactly contained in $\Omega \subset \mathbb{R}^n$, it is well-known that there exists a function $\psi \in C_c^\infty(\Omega)$, such that, $0 \leq \psi \leq 1$ on Ω' via mollification.

Finally, we define the smooth, one-sided, deformation ϕ^ε of ϕ as follows:

$$\phi^\varepsilon = \phi + \rho_+^\varepsilon,$$

where ρ_+^ε is defined as the right side of ρ^ε, that is,

$$\rho_+^\varepsilon(x) = \begin{cases} 1, & 1/2 \leq x_1 \leq 1/2 + \varepsilon/4,\ 0 \leq x_2 \leq x_1, \\ a(x), & 1/2 + \varepsilon/4 \leq x_1 \leq 1/2 + 3\varepsilon/4,\ 0 \leq x_2 \leq x_1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.27}$$

In this definition, $a(x)$ is a smooth function whose values are in $(0, 1)$ for all x.

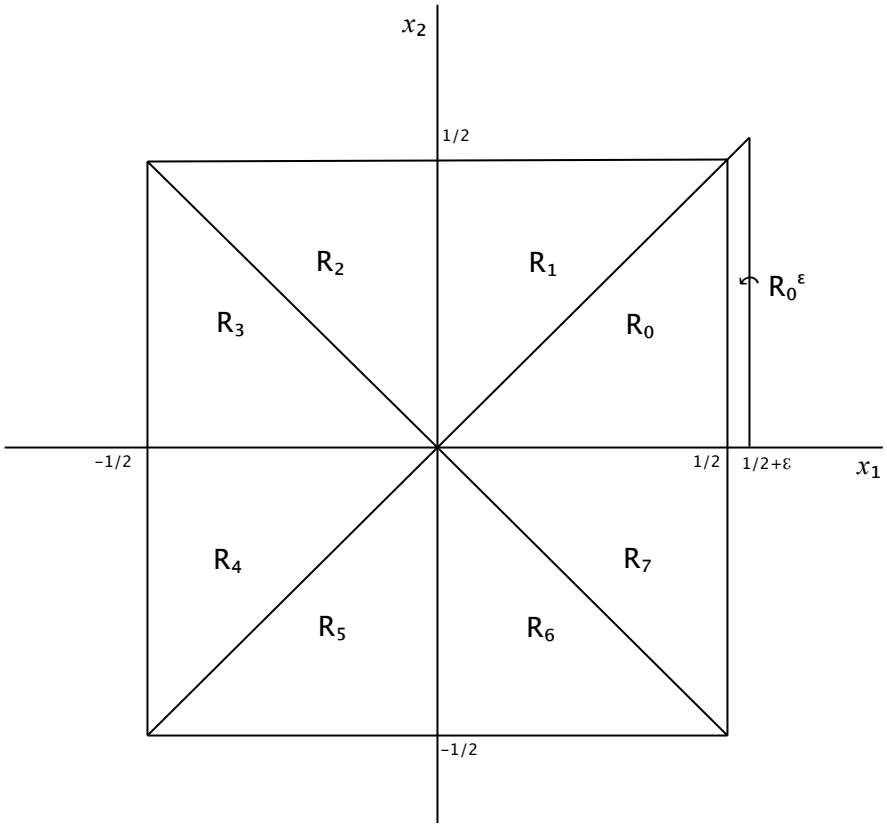


Figure 2.2: Domain of Haar scaling function ϕ^ε smoothed on one side and extended via mollification on R_0^ε.

Proposition 2.18. *The system* $\Phi_{\mathbf{G}}^\varepsilon = \{D_{B_i}T_k\phi^\varepsilon : B_i \in \mathbf{G}, k \in \mathbb{Z}^2\}$ *is a frame for* V_0*, i.e., if* $f \in V_0$*, then there exist constant* C_L*,* $C_U > 0$ *such that*

$$C_L\|f\|^2_{L^2(\mathbb{R}^2)} \leq \sum_{k\in\mathbb{Z}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi^\varepsilon\rangle|^2 \leq C_U\|f\|^2_{L^2(\mathbb{R}^2)}. \tag{2.28}$$

Proof. We start with the upper bound:

$$\begin{aligned}
\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi^\varepsilon\rangle|^2 &= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\langle f,\, D_{B_i}T_k\phi\rangle + \left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right|^2 \\
&= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\Big\{|\langle f,\, D_{B_i}T_k\phi\rangle|^2 + 2\,\langle f,\, D_{B_i}T_k\phi\rangle\left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle \\
&\quad + \left|\left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right|^2\Big\}.
\end{aligned}$$

Using the inequality $2ab \leq a^2 + b^2$ for the middle term, we obtain

$$\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi^{\varepsilon}\rangle|^2 \leq 2\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left\{|\langle f,\, D_{B_i}T_k\phi\rangle|^2 + \left|\left\langle f,\, D_{B_i}T_k\rho_+^{\varepsilon}\right\rangle\right|^2\right\}.$$

Since by Proposition 2.15(i), $\Phi_{\mathbf{G}}$ is an orthonormal basis for V_0, we have

$$\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi\rangle|^2 = \|f\|^2_{L^2(\mathbb{R}^2)},$$

so, we focus on finding a bound for the term

$$\begin{aligned}\mathcal{R}\,(f,\,\varepsilon) &= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\left\langle f,\, D_{B_i}T_k\rho_+^{\varepsilon}\right\rangle\right|^2 \\ &= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\int_{\mathbb{R}^2} f(x)\rho_+^{\varepsilon}\left(B_i^{-1}x-k\right)dx\right|^2.\end{aligned}$$

Let $y = B_i^{-1}x - k$. By definition, $\rho_+^{\varepsilon}(x) = 0$ on $\mathbb{R}^2$ except for $x \in R_0^{\varepsilon} := \{(x_1,\, x_2) : 1 \leq x_1 \leq 1/2+\varepsilon, 0 \leq x_2 \leq x_1\}$, as well as $R_i^{\varepsilon} = B_i R_0^{\varepsilon}$, for $i = 1,\ldots,7$. Note that $x = B_i^{-i}y + B_i^{-1}k$, which we will write as $x = B_i^{-1}y - k$ since B_i is an invertible matrix and $k \in \mathbb{Z}^2$.

Thus, we have

$$\begin{aligned}\mathcal{R}\,(f,\,\varepsilon) &= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\int_{R_i^{\varepsilon}}\rho_+^{\varepsilon}(y)f\left(B_i^{-1}y-k\right)\,\mathrm{d}y\right|^2 \\ &\leq \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left(\int_{R_i^{\varepsilon}}\left|\rho_+^{\varepsilon}(y)\right|D_{B_i}T_k|f(y)|\,\mathrm{d}y\right)^2 \\ &\leq \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\|\rho_+^{\varepsilon}\|_{L^{\infty}\left(R_i^{\varepsilon}\right)}\left(\int_{R_i^{\varepsilon}}D_{B_i}T_k|f(y)|\,\mathrm{d}y\right)^2 \\ &= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left(\int_{R_i^{\varepsilon}}D_{B_i}T_k|f(y)|\,\mathrm{d}y\right)^2,\end{aligned}$$

since, by definition $0 \leq \rho_+^\varepsilon \leq 1$ on R_i^ε. Note that we can write

$$\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left(\int_{R_i^\varepsilon} D_{B_i}T_k|f(y)|\,\mathrm{d}y\right)^2 = \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left(\int_{R_i^\varepsilon} D_{B_i}T_k|f(y)|\chi_{R_i^\varepsilon}(y)\mathrm{d}y\right)^2.$$

Therefore,

$$\begin{aligned}
\mathcal{R}\,(f,\,\varepsilon) &\leq \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left(\int_{R_i^\varepsilon} D_{B_i}T_k|f(y)|\chi_{R_i^\varepsilon}(y)\mathrm{d}y\right)^2 &(2.29)\\
&= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left(\int_{R_i^\varepsilon} |f(y)|D_{B_i}T_k\chi_{R_i^\varepsilon}(y)\,\mathrm{d}y\right)^2\\
&= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\left\langle f,\, D_{B_i}T_k\chi_{R_i^\varepsilon}\right\rangle\right|^2. &(2.30)
\end{aligned}$$

At this stage, it is useful to look at the meaning of (2.29) geometrically, on the fundamental domain. On R_0, $\mathcal{R}\,(f,\,\varepsilon)$ is integrated on translates and images of R_i^ε which overlaps on corners of $[-1/2,\,1/2)^2$ as shown on Figure 2.3. On all the other R_i, we have the contribution of at most 2 other translates of R_i^ε. Therefore, we are integrating most $2|f|$ on each R_i, $i = 0,\ldots,7$, and their translates. Thus,

$$\begin{aligned}
\mathcal{R}\,(f,\,\varepsilon) &= \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle 2f,\, D_{B_i}T_k\phi\rangle|^2\\
&\leq 4\|f\|^2_{L^2(\mathbb{R}^2)}, &(2.31)
\end{aligned}$$

using Proposition 2.15.

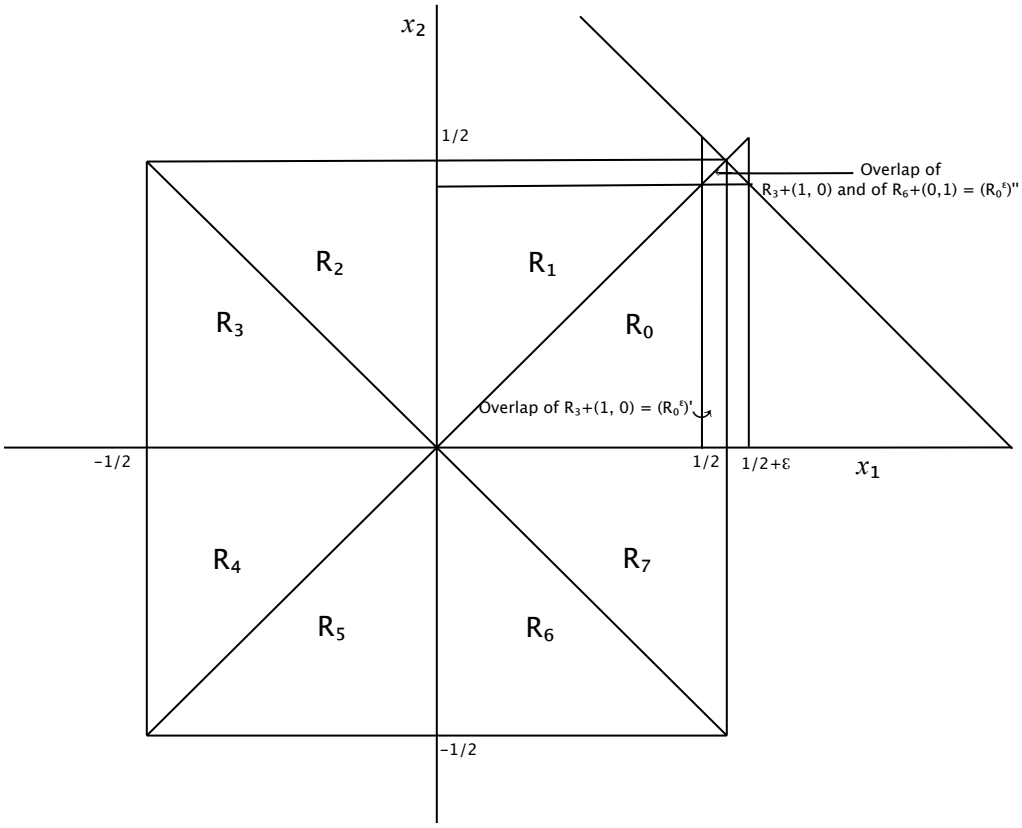


Figure 2.3: The region R_0^ε and its images R_i^ε overlap on $[-1/2, 1/2)^2$.

Next, we compute the lower bound. We have

$$\begin{aligned}
\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi^\varepsilon\rangle|^2 =& \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\langle f,\, D_{B_i}T_k\phi\rangle + \left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right|^2 \\
\geq& \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left||\langle f,\, D_{B_i}T_k\phi\rangle| - \left|\left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right|\right|^2 \\
=& \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\Big\{|\langle f,\, D_{B_i}T_k\phi\rangle|^2 - 2|\langle f,\, D_{B_i}T_k\phi\rangle|\left|\left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right| \\
& + \left|\left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right|^2\Big\} \\
\geq& \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi\rangle|^2 - 2\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi\rangle|\left|\left\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\right\rangle\right|.
\end{aligned}$$

Using Cauchy-Schwartz, we rewrite the last term as

$$\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi\rangle|\big|\big\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\big\rangle\big| \;\leq\; \left(\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi\rangle|^2\right)^{1/2}\times \left(\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\big|\big\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\big\rangle\big|^2\right)^{1/2}.$$

Now, using (2.29),

$$\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\big|\big\langle f,\, D_{B_i}T_k\rho_+^\varepsilon\big\rangle\big| =: \mathcal{R}\,(f,\,\varepsilon) \leq \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\big|\big\langle |f(w)|,\, D_{B_i}T_k\chi_{R_i^\varepsilon}\big\rangle\big|^2.$$

Here, we cannot use the crude upper bound found in (2.31) because it will make our lower bound negative. Instead, once again, we look at this sum on each translate of $[-1/2,\,1/2)^2$ and make use of Figure 2.3 for more careful estimates. We obtain

$$\mathcal{R}\,(f,\,\varepsilon) \leq \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\int_{\mathbb{R}^2}|f(x)|D_{B_i}T_k\chi_{(R_i^\varepsilon)'}\,\mathrm{d}x + \int_{\mathbb{R}^2}|f(x)|D_{B_i}T_k\chi_{(R_i^\varepsilon)''}\,\mathrm{d}x\right|^2.$$

Recall that $f(x)$ is constant on R_i, so we can simplify this expression by finding which proportion of $\int_{\mathbb{R}^2}|f(x)|\chi_{R_i}\mathrm{d}x$ the terms $\int_{\mathbb{R}^2}|f(x)|\chi_{(R_i^\varepsilon)'}\,\mathrm{d}x$ and $\int_{\mathbb{R}^2}|f(x)|\chi_{(R_i^\varepsilon)''}\,\mathrm{d}x$ represent. These proportions are given by, respectively,

$$\begin{aligned} p_1 &:= \frac{\mathrm{Area}((R_i^\varepsilon)')}{\mathrm{Area}(R_1)} = \frac{(1/2)\varepsilon/2}{(1/2)^2/2} = 4\varepsilon \\ p_2 &:= \frac{\mathrm{Area}((R_i^\varepsilon)'')}{\mathrm{Area}(R_1)} = \frac{\varepsilon^2/2}{(1/2)^2/2} = 4\varepsilon^2. \end{aligned}$$

This yields

$$\begin{aligned} \mathcal{R}\,(f,\,\varepsilon) &\leq \sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|(p_1+2p_2)\int_{\mathbb{R}^2}|f(x)|D_{B_i}T_k\chi_{R_i}(x)dx\right|^2 \\ &= (8\varepsilon^2+4\varepsilon)^2\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}\left|\int_{\mathbb{R}^2}|f(x)|D_{B_i}T_k\chi_{R_i}(x)dx\right|^2 \\ &= (8\varepsilon^2+4\varepsilon)^2\|f\|_{L^2(\mathbb{R}^2)}, \end{aligned}$$

and hence, we have

$$\begin{aligned}\sum_{k\in\mathbb{R}^2}\sum_{i=0}^{7}|\langle f,\, D_{B_i}T_k\phi^\varepsilon\rangle|^2 &\geq \|f\|^2_{L^2(\mathbb{R}^2)} - 2(8\varepsilon^2+4\varepsilon)^2\|f\|^2_{L^2(\mathbb{R}^2)} \\ &= (1-2(8\varepsilon^2+4\varepsilon)^2)\|f\|^2_{L^2(\mathbb{R}^2)}.\end{aligned}$$

We simply pick ε such to get $C_L = (1-2(8\varepsilon^2+4\varepsilon)^2) > 0$ to obtain a frame. Figure 2.4 shows that such an ε exists. □

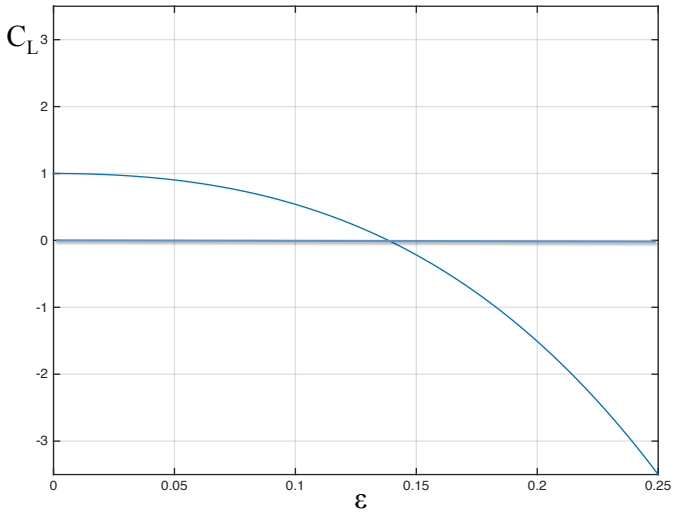


Figure 2.4: This is the plot of the term $\theta(\varepsilon) = (1-2(8\varepsilon^2+4\varepsilon)^2)$ in the lower bound for the approximate $\mathbf{G}_A\mathbf{G}$-MRA frame, $C_L \geq \theta(\varepsilon)\|f\|$, for any $f \in L^2(\mathbb{R}^n)$. If we pick $\varepsilon > 0$ in the interval between 0 and some value around 0.125, we obtain a positive C_L.

Let us summarize our results:

- The system $\Phi^\varepsilon_{\mathbf{G}}$ is a frame for the space V_0.
- Since we have not changed the definition of V_0, the spaces $V_j = A^{-1}V_0$, $j \in \mathbb{Z}$ form a nested sequence for the same reasons as in the orthonormal case.
- The scaling function $\phi(x)$ is "approximately refinable". Indeed,

$$\phi^\varepsilon(x) = \sqrt{2}\left(D_A^{-1}D_{B^{-1}}\phi^\varepsilon(x) + D_A^{-1}D_{B^{-6}}\phi^\varepsilon(x)\right),$$

except for two strips S_1^ε in the middle of $A^{-1}R_1^\varepsilon \cup A^{-1}R_6^\varepsilon + (1/2+\varepsilon/2,\, 1/2+\varepsilon/2)^T$ where $1/4+\varepsilon/2 \le x_1 \le 1/2 \le \varepsilon/2$, and $S_2^\varepsilon = \{1/2+\varepsilon/2 \le x_1 \le 1/2 \le \varepsilon/2,\, 0 \le x_2 \le x_1\}$ which can be made as small as needed by choosing the appropriate ε. This is in part due to the fact that we did not start with a semi-orthogonal frame, i.e., $\mathbb{Z}^2$-translates of the R_i with the added strips are not disjoint.

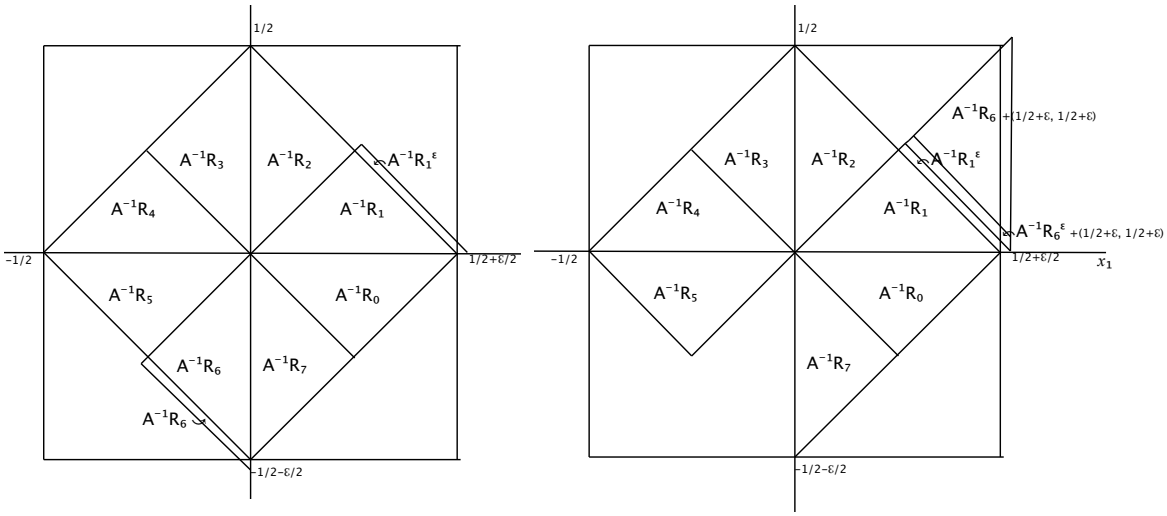


Figure 2.5: No union of images of R_0^ε and their shifts under B_i and A^{-1} gives a cover for R_0^ε. The consequence of this is that ϕ^ε only satisfies an "approximate" $\mathbf{G}_A\mathbf{G}$-MRA condition.

In future work, we would like to make the following modification to our construction. Instead of a one-sided mollification, we could mollify ϕ on all sides. Currently, we have no guarantee that the resulting system is still a frame or a $\mathbf{G}_A\mathbf{G}$-MRA, although the former is likely. However, if it happens to be a frame, we have the added advantage of a smoother function.

2.5 Conclusion

In this chapter, we showed that if we are given the following:

- A subgroup of dilations $\mathbf{G}_A$, formed by integer powers of an invertible, expanding matrix A and a finite subgroup of dilations $\mathbf{G}$, formed by invertible matrices with determinant 1, such that A normalizes $\mathbf{G}$,

- A function $\phi \in L^2(\mathbb{R}^n)$ such that 1) $\Phi_{\mathbf{G}} = \{D_B T_k \phi : B \in \mathbf{G}, k \in \mathbb{Z}^n\}$ is a semi-orthogonal Parseval frame for V_0, a closed subspace of $L^2(\mathbb{R}^n)$ and 2) ϕ satisfies (2.5),

then, we obtain a frame $\mathbf{G}_A\mathbf{G}$-MRA for $L^2(\mathbb{R}^n)$. We also provided an example of an "approximate" $\mathbf{G}_A\mathbf{G}$-MRA.

In the next chapter, we will obtain conditions similar in essence to those listed above for composite diffusion frames for a space $L^2(X, \mu)$ of square-integrable functions on graphs or manifolds. Instead of the condition A normalizes $\mathbf{G}$, we will have the condition that the dilations are "similar" in a specific manner, and we will also start the construction from a frame for a subspace V_0 of $L^2(X, \mu)$.

Chapter 3

Composite Diffusion Wavelet Frames

3.1 Introduction

In this chapter, we give the requirements to construct a Diffusion Frame MRA, Diffusion Wavelet Frames and Composite Diffusion Wavelet Frames for $L^2(X, \mu)$, which contains square-integrable functions defined on a space of homogeneous type X. Subspaces of these frame representations span subspaces from a multiresolution analysis induced by eigenvectors of dyadic powers t_j of an operator S. These powers of S are the discretization of a symmetric diffusion semigroup on $L^2(X, \mu)$.

In [27], R. Coifman and M. Maggioni construct bases of scaling functions approximating the MRA of eigenvectors of S^{t_j} and the corresponding diffusion wavelets using a fast and stable orthogonalization scheme. However, this orthogonalization scheme is complex and makes diffusion wavelets computationally expensive in applications.

Although we will use the idea of diffusion representation systems as described in [25–27], which constitute a Fourier and wavelet analysis for $L^2(X, \mu)$, our goal is to avoid the orthogonalization process completely. Frame representations are an attractive solution for different reasons. One of these reasons is the practicability of frames, e.g., robustness, that we have already discussed in the Introduction and Chapter

2. Moreover, since diffusion frames generalize Coifman and Maggioni's orthogonal system, our construction is appealing on a purely theoretical level. Of course, with frames, we may not have a stable and computationally tractable inversion process existing for orthogonal systems, which allows us to reconstruct the original data. However, since the purpose of obtaining diffusion systems is also to reduce the dimensionality of the data, we can perform the analysis in the embedding space of reduced dimension, and therefore, may never need to reconstruct the original data. In addition, the generalization achieved via composite dilations frames inscribes itself in the framework of constructing systems analogous to those existing on Euclidean spaces. In this case, the goal is to obtain even more flexible systems that would capture properties such as directionality in functions defined on graphs and manifolds.

This chapter is organized as follows. In Section 3.2, we give the notation and definitions used throughout the chapter and state important preliminary results. In Section 3.3, we present the construction of a diffusion frame MRA and diffusion wavelet frames. In Section 3.4, we give sufficient conditions for composite diffusion wavelet frames.

3.2 Background

3.2.1 Notation and Definitions

In this section, we provide the necessary notation and definitions, and also some important results. Many of these can be found in [27], so we will skip proofs of background results that are not essential to the understanding to our construction.

The first definitions are geared towards establishing spaces of homogeneous type. These are abstract spaces that describe many different data domains (graph, manifolds, etc.) of interest here. Much of the genius of the work of Coifman and Maggioni actually lies in viewing the data in the framework of spaces of homogeneous type. We will see in the following that these spaces possess a dyadic structure similar to dyadic cubes on Euclidean spaces, a key property in Coifman and Maggioni's construction, as well as ours.

We start by defining quasi-metric and measure spaces on a set X.

Definition 3.1. The pair $(X,\, d)$, where X is a set and d is a function $d : X \times X \longrightarrow [0, +\infty)$ is a **quasi-metric space** on X, if, for all $x,\, y,\, z \in X$, d satisfies the following properties:

a) Non-negativity: $d(x,\, y) \geq 0$,

b) Identity of indiscernible: $d(x, y) = 0$ if and only if $x = y$,

c) Symmetry: $d(x,\, y) = d(y,\, x)$,

d) Quasi-triangle inequality: there exists $A > 0$ such that

$$d(x,\, y) \leq A\left(d(x,\, z) + d(z,\, y)\right).$$

From this definition, it is clear that the traditional **quasi-metric space** is a metric if A can be set to 1. We denote the open ball of radius δ around $x \in X$ as $B_\delta(x) = \{y \in X : d(x,\, y) < \delta\}$.

Now, we can formally define spaces of homogeneous type:

Definition 3.2. A quasi-metric measure space $(X,\, d,\, \mu)$ with μ, a nonnegative mea-

sure, is a space of **homogeneous type**, if for all $x \in X$, $\delta > 0$, there exists a positive constant C such that

$$\mu\left(B_{2\delta}\left(x\right)\right) \leq C\mu\left(B_{\delta}\left(x\right)\right). \tag{3.1}$$

A measure μ that satisfies (3.1) is said to be C**-doubling**. One simple example of a doubling measure is the Lebesgue measure on a Euclidean space X. Doubling measures are of interest in many areas of analysis because many results from classical harmonic analysis, such as estimations for singular integrals and maximal functions, extend to the general setting of metric spaces when those are equipped with a doubling measure. Here, the doubling measure is a necessary condition to obtain dyadic cubes. Before we define the latter, we make the following assumptions:

(i) X is a connected space;

(ii) For each $\delta > 0$, $x \in X$, $\mu\left(B_{\delta}\left(x\right)\right) < \infty$.

Example 3.3. *Instances of spaces of homogeneous type include $\mathbb{R}^n$ with Euclidean metric and Lebesgue measure, any C^{∞}-compact Riemannian manifold with Riemannian metric and volume, or finite graphs of bounded degree, with shortest path distance and the counting measure [27, 36].*

Now, we can introduce dyadic cubes for X. The following theorem was proved by M. Christ in [22].

Theorem 3.4. *(Dyadic cubes.) Suppose that (X, d, μ) is a space of homogeneous type. Then, there exists a collection of open subsets $\{Q_{\tau}^{j} \subseteq X, j \in \mathbb{Z}, \tau \in \mathcal{T}_j\}$, where*

$\mathcal{T}_j$ is a countable set, and constants $\delta > 1$, $\eta > 0$, and $0 < c_1 \leq c_2 < \infty$, which depend only of A, C, such that:

(i) For each $j \in \mathbb{Z}$, $\mu(X \setminus \bigcup_{\tau \in \mathcal{T}_j} Q_\tau^j) = 0$.

(ii) For each $k \leq j$, $\tau' \in \mathcal{T}_k$, $\tau \in \mathcal{T}_j$, either $Q_{\tau'}^k \subseteq Q_\tau^j$ or $Q_{\tau'}^k \cap Q_\tau^j = \emptyset$.

(iii) For each $k < j$, $\tau \in \mathcal{T}_j$, there exists a unique τ' such that $Q_\tau^j \subseteq Q_{\tau'}^k$.

(iv) Each Q_τ^j contains a dyadic center x_τ^j such that

$$B_{\min\{c_1\delta^j,\, \mathrm{diam(X)}\}}\left(x_\tau^j\right) \subseteq Q_\tau^j \subseteq B_{c_2\delta^j}\left(x_\tau^j\right).$$

(v) For each (l, τ), if $\partial_t Q_\tau^j := \{x \in Q_\tau^j : \rho(x, X \setminus Q_\tau^j) \leq t\delta^j\}$, then $\mu(\partial_t Q_\tau^j) \leq c_2 t^\eta \mu(Q_\tau^j)$.

Here, ρ is the measure of the smallest ball containing x and $X \setminus Q_\tau^j$.

Definition 3.5. a) The collection $\{Q_\tau^j \subseteq X,\, j \in \mathbb{Z},\, \tau \in \mathcal{T}_j\}$ is called a family of **dyadic cubes** for the set X.

b) For each $j \in \mathbb{Z}$, the set $\{Q_\tau^j \subseteq X,\, \tau \in \mathcal{T}_j\}$ is called the set of **dyadic cubes at scale** j and the set $\Gamma_j = \{x_\tau^j\}$ is called the set of **dyadic centers** at scale j.

The meaning of Theorem 3.4 is that spaces of homogeneous type can essentially be understood as Euclidean spaces via the dyadic decomposition. Each of the properties above has an equivalent on $\mathbb{R}^n$. Indeed, part (a) of the theorem states that X has a covering for each dyadic level. Part (b) states that a "cube" in a lower level is either contained in another cube at a higher level or their intersection is empty. Part (c) states that a cube at a lower level must be contained in at most one cube at a higher level. Part (d) states that all sets at some level of the decomposition have roughly the same size, and finally, part (e) states that the area near the boundary of

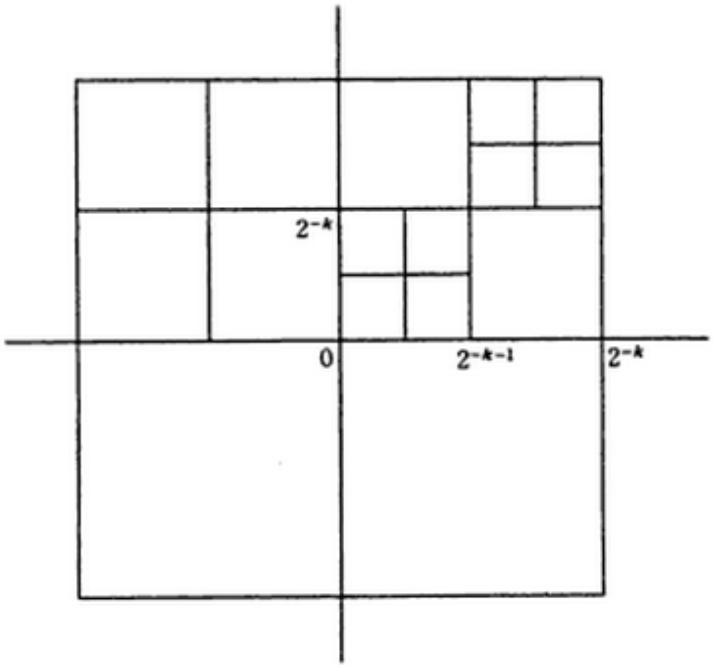


Figure 3.1: Illustration of dyadic cubes for $\mathbb{R}^2$, $\delta = 2$, $\mathcal{T}_j = 2^j\mathbb{Z}^n$.

each cube is small.

Existence of dyadic cubes on Euclidean spaces $\mathbb{R}^n$ are an obvious example of Theorem 3.4, with $\delta = 2$, $\eta = 1$, $c_1 = c_2 = 1$, and $\mathcal{T}_j = 2^j\mathbb{Z}$. An illustration for $\mathbb{R}^2$ is provided in Figure 3.1.

The next definition describes the support set of families for some functions.

Definition 3.6. a) A sequence of functions $\Phi = \{\phi_k\}_{k\in \mathrm{K}}$ defined on X has a **center set**, denoted center$(\Phi) := \{x_k\}_{k\in \mathrm{K}} \subseteq X$, if there exists $\eta > 0$ such that, for each $k \in \mathrm{K}$, we have supp $(\phi_k) \subseteq B_\eta\,(x_k)$.

b) A family of functions with a center set as defined above is said to be η**-local**.

Next, we introduce the notion of ε-span, which allows us to define in which sense a subspace is an approximation for another subspace.

Definition 3.7. a) Let $\mathcal{H}$ is a Hilbert space and let P_V be the orthogonal projection onto V, a closed linear subspace of $\mathcal{H}$. The set of vectors $\{\xi_k\}_{k\in\mathrm{K}}$ **ε-spans** a sequence $\{\phi_j\}_{j\in\mathcal{J}} \in \mathcal{H}$ if, for every $j \in \mathcal{J}$,

$$\|P_{\langle\{\xi_k\}_{k\in\mathrm{K}}\rangle}\phi_j - \phi_j\|_{\mathcal{H}} \leq \varepsilon. \tag{3.2}$$

b) We define the **ε-dimension** of $\{\phi_j\}_{j\in\mathcal{J}}$ by

$$\dim_\varepsilon\left(\{\phi_j\}_{j\in\mathcal{J}}\right) = \inf\left\{\dim(V') : V' \text{ is an } \varepsilon\text{-span of } \left\langle\{\phi_j\}_{j\in\mathcal{J}}\right\rangle\right\}$$

One remark is in order. Terminology-wise, just as in [27] we will abuse notation and say that $\left\langle\{\xi_k\}_{k\in\mathrm{K}}\right\rangle$ is an ε-span of $\left\langle\{\phi_k\}_{k\in\mathrm{K}}\right\rangle$ if Equation (3.2) holds. This is an abuse of notation because a strict application of the definition implies that only set of vectors can ε-span another set of vectors, i.e., ε-spanning is not a relation between subspaces as the latter notation implies.

Example 3.8. *It is useful to look at a simple example to illustrate the significance of Definition 3.7. Suppose that $\{\phi_1, \phi_2, \phi_3\}$ are an orthonormal set. Then, for fixed $\varepsilon > 0$, the set $\{\phi_1, \phi_2\}$ is an ε-span for $\{\phi_1, \phi_2, \varepsilon\phi_3\}$, but not for the original set. Note that any f in the span of $\{\phi_1, \phi_2\}$ can be approximated by $\{\phi_1, \phi_2, \varepsilon\phi_3\}$ with a precision depending on the choice ε. This could be useful in cases where it is more convenient to find an ε-span of a set of vectors than the set of vectors itself.*

Definition 3.9. Let L be a compact, self-adjoint, bounded operator on $L^2(X, \mu)$, with spectrum $\sigma(L)$, and spectral decomposition

$$L = \int_{\sigma(L)} \lambda \, \mathrm{d}E_\lambda.$$

Let $\varepsilon > 0$ be fixed. We define the **ε-approximation** of L, L_ε as follows:

$$L_\varepsilon = \int_{\sigma_\varepsilon(L)=\{\lambda\in\sigma(L):\,|\lambda|>\varepsilon\}} \lambda \,\mathrm{d}E_\lambda. \tag{3.3}$$

L_ε acts on a function $f \in L^2(X,\mu)$ by projecting f onto eigenfunctions of L corresponding to eigenvalues larger than a given precision.

Example 3.10. *One can tie this definition with the concept of ε-span by letting L be an orthogonal projection onto $\{\phi_1, \phi_2, \phi_3\}$ with spectrum $\{1, 1, 0\}$. With $\varepsilon > 0$, the operator L_ε would produce the subspace $\{\phi_1, \phi_2\}$, an ε-span for $\{\phi_1, \phi_2, \varepsilon\phi_3\}$. This concept of an ε-approximation of an operator is key, as the approximation spaces constructed in [27] and in our work are spans of functions obtained by applying powers of an ε-approximation of a family symmetric diffusion semigroup operators, defined in 3.2.2.*

Next, we discuss the construction of Fourier and wavelet bases on graphs and manifolds. The Hilbert space that we consider is the space of square integrable functions $L^2(X,\mu)$ on a space of homogeneous type X. We begin by defining families of operators on $L^2(X,\mu)$.

3.2.2 Symmetric Diffusion Semigroups

The theory of semigroups of operators was established by E. Hille and K. Yoshida in the 1940's and since then, has developed into a beautiful abstract theory with many applications. For the set-up and many of the results used here, we relied on [35, 86, 93]. For an introduction of semigroup operators in the context of PDEs, please, refer to [47]. For the general theory, *Semigroups of Linear Operators and Applications to*

Partial Differential Equations [86], is an acclaimed reference. A. Pazy wrote this textbook at the University of Maryland in 1972-1973, and he did an excellent job in presenting both the abstract theory and basic applications of semigroup operators in a clear, detailed, and stimulating manner. The strength of his contribution is that it establishes the relation between the semigroup and its infinitesimal generator, which is, as we will see in this section, fundamental. Most of the spectral theory results that we use in this section are taken from Pazy's textbook [86].

We begin by defining strongly continuous semigroups of bounded linear operators on a general Banach space $\mathscr{B}$.

Definition 3.11. Let $\mathscr{B}$ be a Banach space. A family of operators $\{S^t\}_{t\geq 0}$, $0 \leq t < \infty$, of bounded linear operators from $\mathscr{B}$ to $\mathscr{B}$ is a **semigroup** of bounded linear operators on X if it satisfies the following:

a) Identity: $S^0 = I$.

b) Semigroup property: $S^{t_1}S^{t_2} = S^{t_1+t_2}$.

Moreover, $\{S^t\}_{t\geq 0}$ is a **strongly continuous** semigroup, denoted C_0 semigroup, if, for all $f \in \mathscr{B}$,

$$\lim_{t\longrightarrow 0} S^t f = f \text{ in } \mathscr{B}. \tag{3.4}$$

Definition 3.12. The linear operator A defined by

$$D(A) = \left\{ f \in \mathscr{B} : \lim_{t\longrightarrow 0} \frac{S^t f - f}{t} \text{ exists} \right\} \tag{3.5}$$

and

$$Af = \lim_{t \longrightarrow 0} \frac{S^t f - f}{t} = \left. \frac{d^+ S^t f}{dt} \right|_{t=0} \text{ for } f \in D(A) \tag{3.6}$$

is the **infinitesimal generator** of the semigroup $\{S^t\}_{t \geq 0}$, $D(A)$ is the domain of A.

The next theorem gives an exponential formula for the infinitesimal generator of a C_0 semigroup.

Theorem 3.13. *(Theorem 8.3, Chap 1, [86].) Let $\{S^t\}_{t \geq 0}$ be a C_0 semigroup on $\mathscr{B}$. If A is the infinitesimal generator of $\{S^t\}_{t \geq 0}$, then*

$$S^t = e^{At}, \tag{3.7}$$

in the sense that, for all $f \in \mathscr{B}$,

$$S^t f = \lim_{n \longrightarrow \infty} \left(I - \frac{t}{n} A \right)^{-n}, \tag{3.8}$$

and this limit is uniform in t on any bounded interval.

In this first, general, definition, t play the role of an index for a set of operators with the semigroup structure. But in the construction that we will propose, the superscript will take the familiar meaning of power of S. Next, we define the family of operators used to build our multiresolution analysis.

Definition 3.14. A strongly continuous semigroup of bounded operators $\{S^t\}_{t \geq 0}$ on $L^2(X, \mu)$ is a **symmetric diffusion semigroup** (or a **Markovian semigroup**) if the following properties hold:

a) Contraction: $\|S^t\|_p \leq 1$ for every $1 \leq p \leq +\infty$.

b) Symmetry: each S^t is self-adjoint.

c) Positivity: for each smooth $f \geq 0$ in $L^2(X, \mu)$, $S^t f \geq 0$.

d) Conservation: $S^t \mathbf{1} = \mathbf{1}$.

Definition 3.16 was first introduced by Stein in [93]. His goal was to obtain classical Fourier analysis tools such as the Littlewood-Paley maximal function, the Littlewood-Paley inequalities and other important results in a very general setting. Symmetric diffusion semigroups provided this setting. In view of Theorem 3.13, it is clear that, by assumption, symmetric diffusion semigroups have an infinitesimal generator for which the exponential formula holds. We give a few examples of symmetric semigroups. To illustrate symmetric diffusion semigroups of operators, we pick two examples from an extensive list provided in [27, 93]. The first example is a typical example of symmetric diffusion semigroup in the context of Partial Differential Equations and the second example illustrates the practical scenarios that we are interested in investigating.

Example 3.15. *a) The semigroup $S^t = e^{tL}$ generated as follows on the interval (a, b):*

$$Lf = a_2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x) + a_1(x)\frac{\mathrm{d}}{\mathrm{d}x}f(x) + a_0(x)f(x),$$

with $a_2(x) > 0$, $a_0(x) \leq 0$, acting on a subspace of $L^2\left((a, b), q(x)\,\mathrm{d}x\right)$, where q is an appropriate weight function, given by imposing appropriate boundary conditions, so that L is formally self-adjoint and generates a semigroup satisfying conditions (a) through (c) in Definition 3.16. If we set $a_0(x) = 0$, with appropriate boundary conditions, then (d) holds as well [93].

b) The semigroup generated by the normalized graph Laplacian. Consider the undirected weighted graph $X = (V, E, W)$, where $V = \{x_1, \ldots, x_i, \ldots\}$ is the set of vertices, $E = \{e\}$ is the set of edges, and $W = \{w_e\}$ is the set of weights assigned to the edges. First define the quasi-metric

$$d(x_i, x_j) = \inf_{\gamma_{x_i, x_j}} \sum_{e \in \gamma_{x_i, x_j}} w_e,$$

where γ_{x_i, x_j} is a path connecting x_i and x_j. Let the diagonal matrix D be defined as

$$D_i = \sum_{\substack{x_j \\ x_i \sim x_j}} w_e,$$

where $\sim$ means that there is an edge between x_i and x_j. Then the normalized Laplacian $L = I - D^{-1/2} W D^{-1/2}$ is a contraction on $L^p(G, \mu_G)$, $\mu_G(i) = D_i$ and is self-adjoint. The normalized Laplacian induces a symmetric diffusion semigroup on $L^2(G, \mu_G)$.

To obtain a convenient spectral theory for our diffusion semigroup, we can add an additional assumption of compactness.

Definition 3.16. A diffusion semigroup $\{S^t\}_{t \geq 0}$ is called **compact** if S^t is compact for $0 < t < \infty$.

Remark 3.17. Note that the notation for compact symmetric diffusion semigroup leaves something to be desired. Indeed, when we write $\{S^t\}_{t \geq 0}$ is **compact**, the notation might suggest that the identity, at $t = 0$ is also compact, but this is not implied by the definition. The next remark expands on why we want to avoid a compact identity operator.

Remark 3.18. In Coifman and Maggioni's paper, the above definition is stated for $t \geq 0$, which means that the identity is compact and hence, that they are only dealing with finite dimensional spaces $L^2(X, \mu)$. Although we also end up constructing finite dimensional multiresolution scaling function spaces, we prefer to adopt Definition 3.16 to keep the possibility of an infinite dimensional space in the corresponding wavelet construction.

We proceed with the spectral theory for $\{S^t\}_{t \geq 0}$.

3.2.3 Spectral Theory for Symmetric Diffusion Semigroups

The next three theorems come from [86]. First, recall the following definitions from function analysis.

Definition 3.19. Let S be a bounded linear operator acting on a Banach space $\mathscr{B}$ over the scalar field $\mathbb{C}$. A complex number λ is in the **spectrum** of S, denoted $\sigma(S)$, if the operator $S - \lambda I$ does not have an inverse.

Now, by the open mapping theorem, if $S - \lambda I$ is one-to-one, then its inverse is bounded. Hence, $\lambda \in \sigma(S)$ if and only if $S - \lambda I$ is either not one-to-one or, not onto. In this occurrence, we can have three distinct sections of the spectrum of S, defined next.

Definition 3.20. a) The value λ is said to be in the **point spectrum** of S, denoted $\sigma_p(S)$, if $S - \lambda I$ is not injective. Then there exist a non-zero function ξ such that $S\xi = \lambda\xi$. Therefore, λ in the points spectrum are eigenvalues of S in the traditional sense of linear algebra.

b) The value λ is said to be in the **continuous spectrum** of S, denoted $\sigma_c(T)$, if $S - \lambda I$ is injective and its range is a dense subset of $\mathscr{B}$, but not the whole $\mathscr{B}$.

c) The value λ is said to be in the **residual spectrum** of S, denoted $\sigma_r(T)$, in $S - \lambda I$ is injective but does not have dense range.

In the next theorem, we characterize the point spectrum (or equivalently, eigenvalues) of a strongly continuous semigroup with infinitesimal generator A.

Proposition 3.21. *(Theorem 2.4, Chap 2, [86].) Suppose $\{S^t\}_{t\geq 0}$ is a strongly continuous semigroup with infinitesimal generator A. If $\sigma_p(A)$ is the point spectrum of A, then*

$$e^{\sigma_p(A)t} \subset \sigma_p(S^t) \subset e^{\sigma_p(At)} \cup \{0\}\,. \tag{3.9}$$

When we add the assumption of compactness of the diffusion semigroup, we have the following:

Proposition 3.22. *(Corollary 3.7, Chap 2, [86].) Suppose $\{S^t\}_{t\geq 0}$ is a compact, strongly continuous semigroup with infinitesimal generator A. For every $-\infty < \alpha \leq \beta < \infty$, the intersection of $\sigma(A)$ with the strip $\alpha \leq \mathrm{Re}(\mu) \leq \beta$, where $\mu \in \sigma(A)$ is compact and contains at most a finite number of eigenvalues.*

Proposition 3.21 has one immediate consequence: if $\{S^t\}_{t\geq 0}$ is compact, then the spectrum $\sigma(A)$ contains only eigenvalues, i.e., $\sigma(A) = \sigma_p(A)$ [86]. Even more important is the consequence of the spectrum of symmetric diffusion semigroup with positive generator. An observation of B. Manning during a conversation on symmetric diffusion semigroups lead to the following theorem:

Theorem 3.23. *Suppose $\{S^t\}_{t\geq 0}$ is a compact, symmetric diffusion semigroup with positive self-adjoint infinitesimal generator A. Then each S^t has at most two eigenvalues: 1 and 0.*

Proof. Since A is positive, we can choose $\alpha = 0$ in Proposition 3.21. Now, by Proposition 3.21 (and $\sigma(A) = \sigma_p(A)$), we know that $e^{\sigma_p(A)t} \subset \sigma(S^t) \subset e^{\sigma(A)t} \cup \{0\}$. Since S^t is a contraction for any t, this implies that the largest that $e^{\mu t}$, for $\mu \in \sigma(A)$ is 1, i.e., we can choose $\beta = 0$ in Proposition 3.21. Therefore, $\sigma(A)$ contains at most one value: 0, and consequently, $\sigma(S^t)$ for any t contains at most two values, 0 and 1. □

Remark 3.24. Based on Theorem 3.23, the spectra of the compact symmetric diffusion semigroup operators with positive generators considered by Coifman and Maggioni (Definition 13, [27]) have at most 1 and 0 for eigenvalues. If 1 is the only eigenvalue, then each S^t is the identity, so $\{S^t\}_{t\geq 0}$ loses practical appeal. Now, suppose that S^t has two eigenvalues 0 and 1. In their construction, Coifman and Maggioni only consider a discrete subset of $\{S^t\}_{t\geq 0}$ indexed by dyadic integers. The consequence of this choice is that their approximation spaces are constant, i.e., $V_0 = V_1 = \ldots$, so there is no actual multiresolution since all $\lambda^{t_j} = 1$. Since we also only consider an integer discretization of $\{S^t\}_{t\geq 0}$ in our construction, we will avoid a trivial construction by assuming that our diffusion semigroup has unbounded negative infinitesimal generator.

Remark 3.25. In all that follows, we will assume that $\{S^t\}_{t\geq 0}$ is a compact diffusion semigroup with unbounded negative generator.

Next, we establish the spectral theory for the discretization of a compact dif-

fusion semigroup $\{S^t\}_{t\geq 0}$. First, consider the spectral theorem for general compact, self-adjoint operators.

Theorem 3.26. *(Spectral theorem for compact, self-adjoint operators [14].) Let $T : L^2(X,\mu) \longrightarrow L^2(X,\mu)$ be a compact, self-adjoint operator. Then there exists a countable orthonormal basis of $L^2(X,\mu)$ consisting of eigenvectors of T.*

Based on this theorem, the eigenvectors associated with each S^t in a compact diffusion semigroup form an orthonormal basis for $L^2(X,\mu)$. The next theorem highlights the fact that the S^t and the infinitesimal generator A share the same eigenvectors.

Theorem 3.27. *(Spectral theorem for compact diffusion semigroups.) Let $\{S^t\}_{t\geq 0}$ be a compact diffusion semigroup $\{S^t\}_{t\geq 0}$ with self-adjoint, negative generator A. Furthermore, suppose that each eigenvalue of $\{S^t\}_{t\geq 0}$ have multiplicity 1. Then, the following hold:*

(i) The eigenvalues μ of A are real and non-zero eigenvalues of each S^t have the form $e^{\mu t}$.

(ii) An eigenvector ξ_μ of A with corresponding eigenvalue $\mu > -\infty$ is also an eigenvector for each operator S^t with corresponding eigenvalue $\lambda^t = e^{\mu t}$.

(iii) An eigenvector ξ_μ of S^t with corresponding eigenvalue $e^{\mu t}$ is also an eigenvector for A with corresponding eigenvalue μ.

(iv) Let ξ_μ be eigenvectors of A, for each $f \in L^2(X,\mu)$, we can write

$$S^t f = \sum_{\mu \in \sigma^*(A)} e^{\mu t} \langle f, \xi_\mu \rangle \xi_\mu, \tag{3.10}$$

where $\sigma^*(A)$ is the spectrum of A without the zero eigenvalues.

For the proof, we need the following result from [86].

Proposition 3.28. *Let $\{S^t\}_{t\geq 0}$ be a C_0-semigroup and let A be its infinitesimal generator. Then,*

(i) ([86], Corollary 1.4, Chap 1.) For $f \in L^2(X,\mu)$, if we define

$$B_\gamma(t)f := \int_0^t e^{\gamma(t-\tau)} S^\tau f \mathrm{d}\tau,$$

we have

$$e^{\gamma t} f - S^t f = B_\gamma(t)(\gamma I - A)f. \tag{3.11}$$

Proof. (*Theorem 3.27*)

(i) The first part of the statement is a standard result for self-adjoint operator on Banach spaces. The second part of the statement follows directly from Proposition 3.21 and the fact that for self-adjoint operators, the point spectrum corresponds to the set of eigenvalues .

(ii) Let $\mu \in \sigma(A)$ and assume that the corresponding eigenvector is ξ_μ, i.e.,

$$A\xi_\mu = \mu\xi_\mu \Longleftrightarrow (I\mu - A)\xi_\mu = 0 \tag{3.12}$$

Then, using Proposition 3.28(i), we have

$$\begin{aligned} S^t\xi_\mu - e^{\mu t}\xi_\mu &= e^{\mu t}\int_0^t e^{-\mu\tau} S^\tau (I\mu - A)\xi_\mu \mathrm{d}\tau \\ &= 0, \end{aligned}$$

i.e., ξ_μ is an eigenvector of S^t.

(iii) Let $\xi_{\mu,1}$ be an eigenvector of S^t with eigenvalue $e^{\mu t}$, where μ is an eigenvalue of

A. We can write

$$S^t \xi_{\mu,1} = e^{\mu t} \xi_{\mu,1}.$$

Suppose that $\xi_{\mu,1}$ is not an eigenvector of A, and let $\xi_{\mu,2}$ be the eigenvector of A corresponding to μ. By part (ii), $\xi_{\mu,2}$ is also an eigenvector of S^t corresponding to $e^{\mu t}$. However, this contradicts the assumption that all eigenvalues of S^t have multiplicity 1. Therefore, the eigenvector ξ_μ of S^t associated with $e^{\mu t}$ is an eigenvector of A with eigenvalue μ.

(iv) Equation (3.10) comes directly from the Spectral Theorem 3.26. Since zero-eigenvalues and the corresponding eigenvectors do not contribute to the sum, they need not appear in the expression.

□

Remark 3.29. Theorem 3.27 is key in our work. It means that the compact symmetric diffusion semigroup operators share eigenvectors with their infinitesimal generators. This will allow us to define a MRA on $L^2(X, \mu)$, which can be approximated using projections of S^t onto certain approximation functions. Moreover, (3.10) will make our computations clear and precise.

3.2.4 Multiresolution Analysis Induced by Symmetric Diffusion Semigroups

Let $\{S^t\}_{t\geq 0}$ be symmetric diffusion semigroup on a space of homogeneous type X. Consider a discretization of $\{S^t\}_{t\geq 0}$ at times $t_j = 2^{j+1} - 1$. Then, by compactness

and Theorem 3.27, for each $f \in L^2(X, \mu)$, we can write

$$S^{t_j} f = \sum_{\lambda \in \sigma(S)} \lambda^{t_j} \langle f, \xi_\lambda \rangle \xi_\lambda. \tag{3.13}$$

We regard S^{t_j} as dilation operator on $L^2(X, \mu)$. Let $\varepsilon \in (0, 1)$ be given and let

$$\sigma_{\varepsilon,j}(S) := \left\{ \lambda \in \sigma(S),\, \lambda^{t_j} \geq \varepsilon \right\}.$$

The subspaces of scaling functions ξ_λ on $L^2(X, \mu)$ can be used to define a multiresolution analysis of **finite dimensional** spaces V_j. Indeed, if we let

$$V_{-1} = L^2(X), \tag{3.14}$$

$$V_j^\varepsilon = \langle \{ \xi_\lambda : \lambda \in \sigma_{\varepsilon,j}(S) \} \rangle, \text{ for } j \geq 0, \tag{3.15}$$

then the V_j^ε's form a multiresolution analysis in the sense that

(i) $V_{-1} = L^2(X)$, $\lim_{j \longrightarrow \infty} V_j^\varepsilon = \langle \{ \xi_\lambda : \lambda = 1 \} \rangle$;

(ii) $V_{j+1}^\varepsilon \subseteq V_j^\varepsilon$ for each $j \geq -1$;

(iii) $\{ \xi_\lambda : \lambda^{t_j} \geq \varepsilon \}$ is an orthonormal basis for V_j^ε.

Properties (i) and (ii) follow from the definition. For (ii), one can use the fact that if $j > j' \geq -1$, then

$$\sigma_{\varepsilon,j}(S) \subseteq \sigma_{\varepsilon,j'}(S). \tag{3.16}$$

The corresponding wavelet decomposition is obtained in the usual way. For each $j \geq -1$, define W_j^ε such as

$$V_j^\varepsilon = V_{j+1}^\varepsilon \oplus W_j^\varepsilon. \tag{3.17}$$

Then $L^2(X) = \bigoplus_{j \geq -1} W_j^{\varepsilon}$ is a wavelet decomposition of $L^2(X, \mu)$.

The next assumption is used to control the dimension of the growth of the approximation spaces.

Definition 3.30. Let $\{S^t\}_{t \geq 0}$ be a compact symmetric diffusion semigroup with spectrum $\sigma(S) = \{\lambda_j : \ \lambda_0 \geq \ldots \geq \lambda_j \ldots \geq 0\}$. If there exists a constant $C > 0$ such that, for each $\lambda \in (0, 1)$,

$$| \{\lambda_j \in \sigma(S) : \lambda_j \geq \lambda\} | \leq C \log_2^{\gamma} \frac{1}{\lambda},$$

then we say that $\{S^t\}_{t \geq 0}$ has γ**-strong decay**.

Finally, we define a few terms that describe the action $\{S^t\}_{t \geq 0}$ on functions in $L^2(X, \mu)$.

Definition 3.31. A symmetric diffusion semigroup $\{S^t\}_{t \geq 0}$ is said to act δ**-locally**, for $\delta > 0$, if for each $x \in X$, and each ϕ that is η-local around x, $S\phi$ is $(\delta + \eta)$-local around x.

Recall that Definition 3.6 described sets of functions $\Phi = \{\phi_k\}_{k \in \mathrm{K}}$ that are η-local. Definition 3.31 tells us that a symmetric diffusion semigroup is δ-local if it acts by expanding the radius of the support of each function $\phi \in \Phi$ by δ.

Definition 3.32. A symmetric diffusion semigroup $\{S^t\}_{t \geq 0}$ is **expanding** if, for each smooth f, $\mathrm{supp}(f) \subseteq \mathrm{supp}(S^t f)$.

Now, we have all the ingredients needed to build our diffusion frames multiresolution analysis and composite dilations diffusion frames construction. Before jumping

to our construction, we describe Coifman and Lafon's diffusion maps and Coifman and Maggioni's diffusion wavelets. The reader is free to skip these sections, as it will not affect the understanding of our construction. However, we chose to present these representation systems here because, 1) they constitute logical steps in the progression towards more generalized representation systems for graphs and manifolds, 2) their advantages justify our construction in the same framework, 3) some of the disadvantages motivate a need for more flexible systems such as our frames.

3.2.5 Diffusion Maps

We present the diffusion maps of Coifman and Lafon [26] quite thoroughly for two reasons. First, they can be viewed as a Fourier analysis on graphs and manifolds and, therefore, are theoretically appealing. Second, they can easily be put in the framework of diffusion semigroup and spaces of homogeneous type and provide a good illustration for the type of data sets we wish to study beyond the present dissertation.

In [26], the authors consider a connected data set X and suppose $(X, \mathcal{A}, \mu)$ is a measure space. In addition, assume that a measure of proximity between $x, y \in X$ is given by a kernel K with the following properties:

a) K is symmetric, i.e., $K(x, y) = K(y, x)$,

b) K is positivity-preserving, i.e., $K(x, y) \geq 0$.

Moreover, K is local in the sense that $K(x, y)$ is large if x and y are "close" and $K(x, y)$ "small" for x and y are far apart. The notion of far apart and close depends on the application. The exponentially weighted distance $K(x, y) = e^{(-||x-y||^2/\sigma)}$, $\sigma > 0$,

is an example of a function with such properties.

Using K, one can further define a notion of similarity between two points x and y via the probability of transition from one point to the other. Starting from any graph G as defined by the pairing $G := (X, K)$, one can make a reversible Markov chain using the normalized graph Laplacian construction that follows. First, define a "transition probability" from x to y is defined as

$$p(x, y) = \frac{K(x, y)}{d(x)}, \tag{3.18}$$

where d, which defines a measure of the density of points in the neighborhood of x, is given by

$$d(x) := \int_X K(x, y)\, \mathrm{d}\mu(y). \tag{3.19}$$

Note that the function $p(x, y)$ is positivity preserving since K and d are positive. However, unless $d(x)$ is a constant function of x or satisfies some specific symmetric properties, $p(x, y)$ will not be symmetric. The quantity $p(x, y)$ can be used to view pairwise similarities as edge flows in a Markov random walk on the graph.

The Markov chain defined above satisfies some interesting stochastic (ergodicity, irreducibility). Let

$$\pi(z) = \frac{d(z)}{\int_X d(x)\, \mathrm{d}\mu(x)}$$

be the stationary distribution of the Markov chain and suppose that the kernel $a(x, y)$, given by

$$a(x, y) = \frac{k(x, y)}{\sqrt{\pi(x)}\sqrt{\pi(y)}},$$

is compact. All these assumptions yield a useful spectral properties the operator P associated with the Markov chain. Indeed, P has a spectrum $\sigma(P)$ and eigenfunctions $\{\xi_\lambda\}_{\lambda\in\sigma(P)}$ as in Theorem 3.27. Here, in particular, by ergodicity, we have $1 = \lambda_0 > |\lambda_1| \geq |\lambda_2| \geq \ldots$. Using the orthogonal basis eigendecomposition, we can write

$$P\xi_\lambda = \lambda\xi_\lambda. \tag{3.20}$$

Employing (3.20), for $t \geq 0$, the probability of transition $p_t(x,\, y)$ is from x to y in t time steps is given by

$$p_t(x,\, y) = \sum_{\lambda\in\sigma(P)} \lambda^t \xi_\lambda(x)\xi_\lambda(y). \tag{3.21}$$

Since the transitions in the Markov chain encode the local geometry between neighboring points, via Equation (3.21), X can be analyzed at increasingly larger scale geometry by looking at up to t steps of the random walk [26]. One can organize the data into clusters by defining them as a region from which the probability of escaping is low as time increases. In this way, the parameter t plays both the roles of time and scale parameters. This idea can motivate the definition of dyadic powers of the semigroup as dilation operators for diffusion wavelets.

Next, Coifman and Lafon introduce a metric that measures the closeness of x and y based on the transition probabilities $p_t(x, z)$ and $p_t(y, z)$ for all $z \in X$.

Definition 3.33. The family of **diffusion distances** $\{D_t\}, t \in \mathbb{N}$ from x to y is

$$D_t^2(x,\, y) := \|p_t(x,\, \cdot) - p_t(y,\, \cdot)\|^2 = \int_X (p_t(x,\, z) - p_t(y,\, z))^2 \, \frac{\mathrm{d}\mu(z)}{\pi(z)}. \tag{3.22}$$

The authors make the observation that, for any given $t \geq 0$, D_t defines a distance

function on X, not only because it satisfies the mathematical definition, but in the sense that it provides an "average" of all possible paths between x and y. If the points x and y are connected by many shorts paths on the graph, then the diffusion distance $D_t(x,\, y)$ is small and vice-versa. Note that, since D_t adds up contributions from all paths between x and y, it gives a reliable notion of proximity. This averaging property also makes this distance more robust to noise unlike the geodesic distance, which is more sensitive to perturbations on the data [26].

Using Equation (3.21), we can rewrite D_t as

$$D_t(x,\, y) = \sqrt{\sum_{\lambda \in \sigma(P) \setminus \{\lambda_0\}} \lambda^{2t} \left(\xi_\lambda(x) - \xi_\lambda(y)\right)^2}. \tag{3.23}$$

Note that the eigenfunction corresponding to λ_0 is omitted in this sum since it is a constant. Now, since $\lambda_n \to 0$ and $|\lambda_i| < 1$ for all i, D_t can be approximated with a finite number of terms within an arbitrary degree of accuracy indexed by some arbitrary constant $\delta > 0$. Let $\sigma_{\delta,t}(P) = \{\lambda \in \sigma(P) : |\lambda|^t > \delta |\lambda_1|^t\}$. The approximation of D_t with relative precision δ is given by

$$D_t^\delta = \sqrt{\sum_{\lambda \in \sigma_\delta(P)} \lambda^{2t} \left(\xi_\lambda(x) - \xi_\lambda(y)\right)^2}. \tag{3.24}$$

Definition 3.34. The family of **diffusion maps** $\{\mathcal{E}_t\}_{t \in \mathbb{R}}$ is defined as the set

$$\mathcal{E}_t(x) := \left\{\lambda^t \xi_\lambda(x)\right\}_{\lambda \in \sigma_\delta(S)}, \tag{3.25}$$

where each component $\lambda^t \xi_\lambda(x)$ is called a **diffusion coordinate**.

The mapping $D_t^\delta : X \to \mathbb{R}^{d(\delta,t)}$, where $d(\delta, t) = |\sigma_\delta(P)|$ embeds the data in a Euclidean

space, where diffusion distances are preserved up to a relative accuracy δ, i.e.,

$$D_t(x,\, y) = \|\mathcal{E}_t(x) - \mathcal{E}_t(y)\|_{L^2\left(\mathbb{R}^{d(\delta,\, t)}\right)}.$$

This last equation tells us that, for a given time scale, $\mathcal{E}_t$ embeds the space X in the Euclidean space $\mathbb{R}^{d(\delta,t)}$, where the coordinates of each point x is given by the diffusion coordinate $\lambda^t\xi_\lambda(x)$ and the distance between two points x and y is given $D_t(x,\, y)$ [26]. Therefore, diffusion maps can be viewed as a reparametrization of the data. In this reparametrized space, we can achieve dimension reduction. Indeed, for a fixed t, $d(\delta, t)$ is the dimension of the embedding space, hence, the smaller it is, the better the dimensionality reduction from X to $\mathbb{R}^{d(\delta,t)}$. Now, for a choice of δ, the value $s(\delta,\, t)$ depends on how early the ratio $|\lambda|/|\lambda_1^t|$ exceeds δ. Therefore, the decay of the spectrum governs the dimension of the embedding space. In particular, if P has γ-strong decay,

$$d(\delta, t) \leq Ct^{-\gamma} \log_2^\gamma \left(\frac{1}{\delta}\right).$$

Remark 3.35. Based on this construction, diffusion maps are a Fourier analysis for a graph meeting certain requirements. To put diffusion maps in the framework of symmetric diffusion semigroup/space of homogeneous type, it suffices to note that D_t is a quasi-metric on X (more precisely, it is a metric) and assume that the powers $\{P^t\}_{t\geq 0}$ of P form a compact symmetric diffusion semigroup acting on $L^2(X, \mu)$, where (X, D_t, μ) is a space of homogeneous type.

3.2.6 Diffusion Wavelets

We consider a symmetric diffusion semigroup $\{S^t\}_{t\geq 0}$ on a space of homogeneous type X. Based on the experience with diffusion maps, in which powers of the Markovian operator can be used to define different scales, we regard S as a dilation operator. Suppose that we have the set-up described in 3.2.4. Although the bases for the spaces in Equation (3.17) are simple candidates for diffusion wavelets, they are made of eigenfunctions of S, which are typically highly non-localized, just as in the Euclidean case. Moreover, the computation of the eigenfunctions of S can be computationally costly. To get around this difficulty, Coifman and Maggioni construct ε-approximations of the V_j^ε with localized bases of functions. The justification for the idea is a Heisenberg principle [84] that states that "eigenfunctions have a smoothness or frequency content or scale determined by the corresponding eigenvalues, and can be reconstructed by maximally localized bump functions, or atoms, at that scale" [27].

We briefly describe diffusion wavelets constructed in the setting spaces of homogeneous type:

1. **Multiscale dyadic orthogonalization.** Starting with a set of "bump" functions defined on a space of homogeneous type, using "modified multiscale Gram-Schmidt", construct a set of well-localized orthonormal functions spanning the same subspace, up to a given precision. The dyadic cubes property of spaces of homogeneous type is essential to this step.

2. **Approximate multiresolution scaling functions.** Apply the multiscale dyadic orthogonalization to families of the form $\{S^{t_j}\delta_x\}_{x\in X}$. Here $\{S^t\}_{t\geq 0}$ is a compact

diffusion semigroup with γ-strong decay, t_j a dyadic power, and the δ_x are either a set of Dirac δ-functions in the discrete setting, or a mollification of those in the continuous case. The the multiscale dyadic orthogonalization of $\{S^{t_j}\delta_x\}_{x\in X}$ yields a family of orthonormal bases $\{\Phi_j\}_{j\geq 0}$, which ε-span the V_j^ε, $j \geq 0$.

3. **Diffusion wavelets.** Build bases for the spaces W_j^ε, $j \geq 0$ by applying a the modified multiscale Gram-Schmidt procedure to the set of functions

$$\{(P_j - P_{j+1})\phi_{j,k}\}_{k\in \mathrm{K}_j},$$

where P_j is the projection onto V_j^ε, and $\phi_{j,k} \in \Phi_j$ are functions from the approximate multiresolution analysis.

The procedure modified multiscale Gram-Schmidt procedure as well as the proofs of the validity of the multiscale dyadic orthogonalization and the approximate multiresolution scaling functions construction can be found in detail in [27].

3.3 Frame Multiresolution Analysis and Diffusion Wavelet Frames

3.3.1 Frame Multiresolution Analysis

As stated in the introduction, diffusion wavelets require a computationally expensive orthogonalization process [27]. Although the cost of this process can be significantly improved using the fact that large powers of S can be compressed if S is of low rank, it is still high [27]. This leads us to direct our attention to frames, which would allow us to forego of this orthogonalization process when building the multiresolution of

approximation spaces.

We first construct a diffusion frame multiresolution analysis from which we obtain diffusion wavelets frames. In the next section, we will use some of the results established here to construct composite diffusion frames. We start by defining an operator between the subspaces V_j^ε, $j \geq 0$ and study its properties.

Assumption: In the remaining of this chapter, we make the assumption that all the eigenvalues of S have multiplicity 1.

Notation 3.36. *Let $\{S^t\}_{t\geq 0}$ be a compact symmetric semigroup and let $t_j = 2^{j+1}-1$. For $f \in L^2(X,\mu)$, we denote by $S_{\varepsilon,j}$ the operator given by*

$$S_{\varepsilon,j} f = \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} \langle f, \xi_\lambda \rangle \xi_\lambda, \tag{3.26}$$

Remark 3.37. Notice the difference between the operator $S_{\varepsilon,j}$ and S^{t_j}, $j \geq 0$. Recall that the action of S^{t_j} is given by

$$S^{t_j} f = \sum_{\lambda \in \sigma(S)} \lambda^{t_j} \langle f, \xi_\lambda \rangle \xi_\lambda.$$

To understand this difference, suppose that we have a family of function $\{\phi\}_{k\in \mathrm{K}}$, where K is some indexing set. Then, for $j \geq 0$, $\{S_{\varepsilon,j}\phi\}_{k\in\mathbb{Z}}$ is an ε-span of $\{S^{t_j}\phi\}_{k\in\mathbb{Z}}$.

Proposition 3.38. *Let $\{S^t\}_{t\geq 0}$ be a compact symmetric semigroup and $S_{\varepsilon,j}$ be as in* (3.26)*. Additionally, given $\varepsilon \in (0,\,1)$, $t_j = 2^{j+1}-1$, define the spaces V_j^ε as in 3.15. Then the following are true:*

(i) For each $j \geq 0$, $S_{\varepsilon,j} : V_j^\varepsilon \longrightarrow V_j^\varepsilon$ is a closed range, bounded operator. Moreover, V_j^ε is invariant under $S_{\varepsilon,j}$.

(ii) For each j', $j \geq 0$, and $j' < j$, the operator $S_{\varepsilon,j} : V_{j'}^{\varepsilon} \longrightarrow V_j^{\varepsilon}$ is a closed range, bounded operator from $V_{j'}^{\varepsilon}$ to V_j^{ε}.

(iii) For each $j \geq 0$, $S_{\varepsilon,j}$ is self-adjoint, and hence, so is its pseudoinverse $S_{\varepsilon,j}^{\dagger}$.

(iv) For each $f \in V_j^{\varepsilon}$, we can write

$$S_{\varepsilon,j} f = S_{\varepsilon,j}^{\dagger} (S_{\varepsilon,j})^2 f. \tag{3.27}$$

Proof. (i) The first assertion follows directly from the definition of $S_{\varepsilon,j}$. Indeed, if $f \in V_j^{\varepsilon}$, then f can be written as $f = \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} c_\lambda \xi_\lambda$. However,

$$\begin{aligned} S_{\varepsilon,j} f &= \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} c_\lambda S_{\varepsilon,j} \xi_\lambda \\ &= \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} c_\lambda \left[\sum_{\lambda' \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} \langle \xi_\lambda, \xi_{\lambda'} \rangle \xi_{\lambda'} \right]. \end{aligned}$$

Now, if $\lambda \neq \lambda'$, then $\langle \xi_\lambda, \xi_{\lambda'} \rangle = 0$.

If $\lambda = \lambda'$, then ξ_λ and $\xi_{\lambda'}$ coincide. Thus $\langle \xi_\lambda, \xi_{\lambda'} \rangle = 1$ and

$$S_{\varepsilon,j} f = \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} c_\lambda \xi_\lambda.$$

For the second part of the assertion, the range is closed by linearity and boundedness of $S_{\varepsilon,j}$ follows directly from boundedness of S.

(ii) Let $f \in V_{j'}$. We can write $f = \sum_{\lambda' \in \sigma_{j'}(S)} c_\lambda \xi_{\lambda'}$. Then,

$$\begin{aligned} S_{\varepsilon,j} f &= \sum_{\lambda' \in \sigma_{\varepsilon,j'}(S)} c_{\lambda'} S_{\varepsilon,j} \xi_{\lambda'} \\ &= \sum_{\lambda' \in \sigma_{\varepsilon,j'}(S)} c_{\lambda'} \left[\sum_{\lambda \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} \langle \xi_{\lambda'}, \xi_\lambda \rangle \xi_\lambda \right]. \end{aligned}$$

Note that, since $\sigma_{\varepsilon,j}(S) \subseteq \sigma_{\varepsilon,j'}(S)$, for $\lambda' \in \sigma_{\varepsilon,j}(S)$, we have

$$\sum_{\lambda \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} \langle \xi_{\lambda'}, \xi_\lambda \rangle \xi_\lambda = \lambda^{t_j} \xi_\lambda,$$

and $\lambda' \notin \sigma_{\varepsilon,j}(S)$,

$$\sum_{\lambda \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} \langle \xi_{\lambda'}, \xi_\lambda \rangle \xi_\lambda = 0.$$

This yields

$$S_{\varepsilon,j} f = \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} c_\lambda \xi_\lambda.$$

(iii) Self-adjointness of $S_{\varepsilon,j}$ follows directly from self-adjointness of S. We obtain that $S_{\varepsilon,j}^\dagger$ using Proposition 1.11 (iii):

$$[S_{\varepsilon,j}^\dagger]^* = \left[S_{\varepsilon,j}^*\right]^\dagger = S_{\varepsilon,j}^\dagger.$$

(iv) Using Proposition 1.11 (i), for $f \in V_j^\varepsilon$, we have

$$\begin{aligned} S_{\varepsilon,j} f &= (S_{\varepsilon,j} S_{\varepsilon,j}^\dagger)^* S_{\varepsilon,j} f \\ &= S_{\varepsilon,j}^\dagger S_{\varepsilon,j} S_{\varepsilon,j} f \\ &= S_{\varepsilon,j}^\dagger (S_{\varepsilon,j})^2 f. \end{aligned}$$

□

Theorem 3.39. *(Frame multiresolution analysis.) Suppose that we are given the following:*

a) A bounded space of homogeneous type (X, d, μ),

b) A compact symmetric semigroup $\{S^t\}_{t \geq 0}$ *on* $L^2(X, \mu)$.

c) A precision $\varepsilon \in (0, 1)$, *such that, for* $t_j = 2^{j+1} - 1$, *the* V_j^ε *are defined as in* (3.15),

d) A frame sequence $\Phi_0 = \{\phi_{0,k}\}_{k\in \mathrm{K}}$, $\mathrm{K} < \infty$, *with frame constants* C_1, $C_2 > 0$, *i.e.,*

$$C_1\|f\| \leq \sum_{k\in \mathrm{K}} |\langle f,\, \phi_{0,k}\rangle| \leq C_2\|f\|,$$

such that, $\langle \Phi_0\rangle = V_0^{\varepsilon}$.

For each $j \geq 1$, *define the sequence.*

$$\Phi_j := S_{\varepsilon,j}\Phi_0, \tag{3.28}$$

where S_ε *is defined with as in* (3.26). *Then,* $\{\Phi_j\}_{j\geq 0}$ *has the following properties:*

(i) $\langle \Phi_j\rangle = V_j^{\varepsilon}$ *and hence* $\langle \Phi_{j+1}\rangle \subseteq \langle \Phi_j\rangle$,

(ii) Φ_j *is a frame for* $\langle \Phi_j\rangle$ *with frame constants* $C_1', C_2 > 0$.

Proof. Let

$$\phi_{j,k} = S_{\varepsilon,j}\phi_{0,k}, \quad j \geq 0, k \in \mathrm{K}.$$

(i) For the first assertion, we begin by showing that $V_j^{\varepsilon} \subseteq \langle \Phi_j\rangle$. By definition, we have $\langle \Phi_0\rangle = V_0^{\varepsilon}$. Let $f \in V_j^{\varepsilon}$, $j \geq 1$. Then, we can write

$$f = \sum_{\lambda\in\sigma_{\varepsilon,j}(S)} c_\lambda \xi_\lambda,$$

for some $\{c_\lambda\}_{\lambda\in\sigma_{\varepsilon,j}(S)}$.

But then

$$\begin{aligned} f &= \left(\sum_{\lambda\in\sigma_{\varepsilon,j}(S)} \lambda^{-t_j} c_\lambda \lambda^{t_j}\xi_\lambda\right) \\ &= \left(\sum_{\lambda\in\sigma_{\varepsilon,j}(S)} \lambda^{-t_j} c_\lambda S_{\varepsilon,j}\xi_\lambda\right). \end{aligned}$$

Since Φ_0 is a frame for V_0^ε, we can write $\xi_\lambda = \sum_{k\in \mathrm{K}} c_{\lambda,0,k}\phi_{0,k}$ for some $\{c_{\lambda,0,k}\}_{k\in\mathrm{K}}$. Then,

$$\begin{aligned} f = &= \sum_{\lambda\in\sigma_{\varepsilon,j}(S)} \lambda^{-t_j} c_\lambda S_{\varepsilon,j}\left(\sum_{k\in\mathrm{K}} c_{\lambda,0,k}\phi_{0,k}\right) \\ &= \sum_{\lambda\in\sigma_{\varepsilon,j}(S)}\sum_{k\in\mathrm{K}} \lambda^{-t_j} c_\lambda c_{\lambda,0,k} S_{\varepsilon,j}\phi_{0,k} \\ &= \sum_{k\in\mathrm{K}}\left[\sum_{\lambda\in\sigma_{\varepsilon,j}(S)} \lambda^{-t_j} c_\lambda c_{\lambda,0,k}\right] S_{\varepsilon,j}\phi_{0,k} \\ &= \sum_{k\in\mathrm{K}} c_{j,k}\phi_{j,k}, \end{aligned}$$

where $c_{j,k} = \left[\sum_{\lambda\in\sigma_{\varepsilon,j}(S)} \lambda^{-t_j} c_\lambda c_{\lambda,0,k}\right]$. Note that, since the V_j^ε are finite dimensional, $|\sigma_{\varepsilon,j}(S)| < \infty$, and by assumption $\mathrm{K} < \infty$. Therefore all the sums above are finite, which explains why we were able to switch the summation order and why $c_{j,k}$ is well defined and in $\ell^2(\mathbb{N})$.

The fact that $\langle\Phi_j\rangle \subseteq V_j^\varepsilon$ can be obtained using the definition of Φ_j and a similar argument. If $f \in \langle\Phi_j\rangle$, for $j \geq 1$, there exists $\{c_{j,\,k}\}_{k\in\mathrm{K}} \subseteq \ell^2$ such that we can write

$$\begin{aligned} f &= \sum_{k\in\mathrm{K}} c_{j,k}\phi_{j,k} \\ &= \sum_{k\in\mathrm{K}} c_{j,k} \sum_{\lambda\in\sigma_{\varepsilon,j}(S)} \lambda^{tj}\tilde{c}_\lambda \xi_\lambda \\ &= \sum_{\lambda\in\sigma_{\varepsilon,j}(S)}\left[\sum_{k\in\mathrm{K}} c_{j,k}\lambda^{tj}\tilde{c}_\lambda\right]\xi_\lambda. \end{aligned}$$

(ii) For $f \in \langle\Phi_j\rangle$, we have

$$\sum_{k\in\mathrm{K}} |\,\langle f,\,\phi_{j,k}\rangle\,|^2 = \sum_{k\in\mathrm{K}} |\,\langle f,\,S_{\varepsilon,j}\phi_{0,k}\rangle\,|^2 = \sum_{k\in\mathrm{K}} |\,\langle S_{\varepsilon,j}f,\,\phi_{0,k}\rangle\,|^2,$$

Now, by part (i), we know that $f \in V_j^\varepsilon$. Then, by invariance of V_j^ε under $S_{\varepsilon,j}$, given by Lemma 3.38 (i), we have $S_{\varepsilon,j}f \in V_j^\varepsilon$. This, combined with the fact that the V_j^ε

are nested implies $S_{\varepsilon,j} f \in V_0$. Therefore, we have

$$C_1 \|S_{\varepsilon,j} f\|^2 \leq \sum_{k \in \mathrm{K}} |\langle S_{\varepsilon,j} f, \phi_{0,k} \rangle|^2 \leq C_2 \|S_{\varepsilon,j} f\|^2.$$

It is easy to see that Φ_j is a Bessel sequence:

$$\|S_{\varepsilon,j} f\|^2 \leq \|S_{\varepsilon,j}\|^2 \|f\|^2 \leq \|f\|^2,$$

where, for the second inequality, we have employed the fact that $S_{\varepsilon,j}$ is a contraction. For the lower bound, using Lemma 3.38 (iv), we can write:

$$f = S^{\dagger}_{\varepsilon,j} (S_{\varepsilon,j})^2 f,$$

which implies

$$\|f\| = \|S^{\dagger}_{\varepsilon,j} (S_{\varepsilon,j})^2 f\| \leq \|S^{\dagger}_{\varepsilon,j}\| \|S_{\varepsilon,j}\| \|S_{\varepsilon,j} f\| \leq \|S^{\dagger}_{\varepsilon,j}\| \|S_{\varepsilon,j} f\|.$$

Therefore,

$$\begin{aligned} \sum_{k \in \mathrm{K}} |\langle S_{\varepsilon,j} f, \phi_{0,k} \rangle|^2 &\geq C_1 \|S_{\varepsilon,j} f\|^2 \\ &\geq C_1 \frac{\|f\|^2}{\|S^{\dagger}_{\varepsilon,j}\|^2} = C_1' \|f\|^2. \end{aligned}$$

□

Remark 3.40. Note that Theorem 3.39 as well as the Theorem that will follow 3.41, there is nothing particular about V_0^{ε}. Just as for MRA constructions in Euclidean spaces, we start with a frame for any V_j^{ε} and build approximations of spaces $V_{j'}^{\varepsilon}$ for $j' > j$.

The next result states that if we starts with a frame $\{\phi_k\}_{k \in \mathrm{K}}$ whose span is approximately V_0^{ε}, then $\{S_{\varepsilon,j} \phi_k\}_{k \in \mathrm{K}}$ gives us an approximate multiresolution analysis.

In certain problems, this will relieve us from finding a frame that spans exactly V_0. In particular, we would like to be able to use the spanning functions at any step j of the construction in [27] (minus the orthogonalization process) as the starting frame for some space V_j^ε.

Theorem 3.41. *(Approximate frame multiresolution analysis) Suppose that we have all the assumptions of Theorem 3.39 except that we are given a sequence of functions* $\tilde{\Phi}_0 = \{\phi_{0,i}\}_{i\in\mathcal{I}}$, *which is a* δ*-span for* $\Phi_0 = \{\phi_{0,k}\}_{k\in\mathrm{K}}$, *where* Φ_0 *is a frame for* V_0^ε *with frame constants* C_1, $C_2 > 0$. *Then there exists a sequence* $\left\{\tilde{\Phi}_j\right\}_{j\geq 0}$ *such that for each* $j \geq 0$, $\tilde{\Phi}_j$ δ*-spans* Φ_j, *a frame for* V_j^ε, *with frame bounds* $C_1', C_2' > 0$.

Proof. For $j \geq 1$, $t_j = 2^{j+1} - 1$. Define $\tilde{\Phi}_j = S_{\varepsilon,j}\tilde{\Phi}$, $\Phi_j = S_{\varepsilon,j}\Phi$. Similarly, let $\tilde{\phi}_{j,i} = S_{\varepsilon,j}\tilde{\phi}_{0,i}$, $i \in \mathcal{I}$ and $\phi_{j,k} = S_{\varepsilon,j}\phi_{0,k}$, $k \in \mathrm{K}$.

The fact that we can obtain a frame Φ_j for V_j^ε, with frame bounds C_2 and $C_1' > 0$, $C_1' = \dfrac{C_1}{\|S_{\varepsilon,j}^\dagger\|}$ comes from Theorem 3.39. We want to show that $\tilde{\Phi}_j$ δ-spans Φ_j for $j \geq 0$. Let U be the frame operator for Φ_0.

Employing the formula for the frame operator of for $\{S_{\varepsilon,j}\phi_{0,k}\}_{k\in}$ in Proposition 1.8, we have, for each $\tilde{\phi}_{j,i}$, $i \in \mathcal{I}$,

$$\begin{aligned}\|P_{\langle\Phi_j\rangle}\tilde{\phi}_{j,i} - \tilde{\phi}_{j,i}\| &= \|\sum_{k\in\mathrm{K}}\left\langle\tilde{\phi}_{j,i},\, (S_{\varepsilon,j}US_{\varepsilon,j})^{-1}\phi_{j,k}\right\rangle\phi_{j,k} - \tilde{\phi}_{j,i}\| \\ &= \|\sum_{k\in\mathrm{K}}\left\langle S_{\varepsilon,j}(S_{\varepsilon,j}US_{\varepsilon,j})^{-1}S_{\varepsilon,j}\tilde{\phi}_{0,\,i},\, \phi_{0,k}\right\rangle S_{\varepsilon,j}\phi_{0,k} - \tilde{\phi}_{j,i}\|,\end{aligned}$$

where we have used the fact that $S_{\varepsilon,j}$ is self-adjoint.

Now, by Proposition 1.11 (v), we obtain

$$\|P_{\langle\Phi_j\rangle}\tilde{\phi}_{j,i} - \tilde{\phi}_{j,i}\| = \|\sum_{k\in \mathrm{K}} \left\langle S_{\varepsilon,j}(S_{\varepsilon,j}US_{\varepsilon,j})^{\dagger}S_{\varepsilon,j}\tilde{\phi}_{0,\,i},\, \phi_{0,k}\right\rangle S_{\varepsilon,j}\phi_{0,k} - \tilde{\phi}_{j,i}\|,$$

and by Proposition 3.21 (vi) and subsequent algebraic manipulations, we have

$$\begin{aligned}
\|P_{\langle\Phi_j\rangle}\tilde{\phi}_{j,i} - \tilde{\phi}_{j,i}\| &= \|\sum_{k\in \mathrm{K}} \left\langle S_{\varepsilon,j}(S_{\varepsilon,j})^{\dagger}U^{-1}(S_{\varepsilon,j})^{\dagger}S_{\varepsilon,j}\tilde{\phi}_{0,\,i},\, \phi_{0,k}\right\rangle S_{\varepsilon,j}\phi_{0,k} - \tilde{\phi_{j,\,i}}\| \\
&= \|\sum_{k\in \mathrm{K}} \left\langle U^{-1}\tilde{\phi}_{0,\,i},\, \phi_{0,k}\right\rangle S_{\varepsilon,j}\phi_{0,k} - S_{\varepsilon,j}\tilde{\phi}_{0,i}\| \\
&= \|S_{\varepsilon,j}\left(\sum_{k\in \mathrm{K}} \left\langle U^{-1}\tilde{\phi}_{0,\,i},\, \phi_{0,k}\right\rangle \phi_{0,k} - \tilde{\phi}_{0,\,i}\right)\| \\
&\leq \|S_{\varepsilon,j}\|\|\left(\sum_{k\in \mathrm{K}} \left\langle U^{-1}\tilde{\phi}_{0,\,i},\, \phi_{0,k}\right\rangle \phi_{0,k} - \tilde{\phi}_{0,\,i}\right)\|.
\end{aligned}$$

Using the assumption and $\|S_{\varepsilon,j}\| \leq 1$,

$$\begin{aligned}
\|P_{\langle\Phi_j\rangle}\tilde{\phi}_{j,i} - \tilde{\phi}_{j,i}\| &\leq \|\left(\sum_{k\in \mathrm{K}} \left\langle U^{-1}\tilde{\phi}_{0,1},\, \phi_{0,k}\right\rangle \phi_{0,k} - \tilde{\phi}_{0,\,i}\right)\| \\
&\leq \delta.
\end{aligned}$$

Thus, applying the definition of δ-span, we have that, for each $j \geq 0$, $\tilde{\Phi}_j$ δ-spans Φ_j. □

3.3.2 Diffusion Wavelet Frames

The construction of diffusion wavelet frames follows easily from the previous results and a classical argument from wavelet frames on Euclidean spaces [8, 9]. For $j \geq 0$, recall that we define each space W_j^{ε} as the orthogonal complement of V_{j+1}^{ε} in V_j^{ε}, i.e.,

$$V_j^{\varepsilon} = V_{j+1}^{\varepsilon} \oplus W_j^{\varepsilon}. \tag{3.29}$$

Just as for the frame multiresolution analysis in Theorem 3.39, we define a

particular operator that is invariant on the W_j^ε and has some special properties.

Notation 3.42. *Let $\{S^t\}_{t\geq 0}$ be a compact symmetric semigroup and let $t_j = 2^{j+1}-1$. For $f \in L^2(X, \mu)$, we denote by $S_{\varepsilon,j}^\perp$ the operator as follows:*

$$S_{\varepsilon,j}^\perp f = \sum_{\lambda \in \sigma_{\varepsilon,j}^c(S)} \lambda^{t_j} \langle f, \xi_\lambda \rangle \xi_\lambda, \tag{3.30}$$

where $\sigma_{\varepsilon,j}^c(S) = \sigma_{\varepsilon,j}(S) \backslash \sigma_{j+1}^\varepsilon(S)$.

Proposition 3.43. *Let $\{S^t\}_{t\geq 0}$ be a compact symmetric semigroup and $S_{\varepsilon,j}^\perp$ as in (3.30). Additionally, given $\varepsilon \in (0, 1)$, $t_j = 2^{j+1}-1$, define the spaces V_j^ε as in (3.15) and W_j^ε as in (3.29). Then the following are true:*

(i) For each $j \geq 0$, $S_{\varepsilon,j}^\perp : W_j^\varepsilon \longrightarrow W_j^\varepsilon$ is a closed range, bounded operator. Moreover, W_j^ε is invariant under $S_{\varepsilon,j}^\perp$.

(ii) For each $j \geq 0$, the operator $S_{\varepsilon,j} : V_j^\varepsilon \longrightarrow W_j^\varepsilon$ is closed range, bounded operator from V_j^ε to W_j.

(iii) For each $j \geq 0$, $S_{\varepsilon,j}^\perp$ is self-adjoint, and hence so is $(S_{\varepsilon,j}^\perp)^\dagger$.

(iv) For each $f \in W_j^\varepsilon$, we can write

$$S_{\varepsilon,j}^\perp f = (S_{\varepsilon,j}^\perp)^\dagger (S_{\varepsilon,j}^\perp)^2 f. \tag{3.31}$$

Proof. (i) Let $f \in W_j^\varepsilon$. Then $f = \sum_{\lambda \in \sigma_{\varepsilon,j}^c(S)} c_\lambda \xi_\lambda$. Therefore, using 3.30 and orthonormality of the eigenfunctions, we have

$$\begin{aligned} S_{\varepsilon,j}^\perp f &= \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} c_\lambda \left[\sum_{\lambda' \in \sigma_{\varepsilon,j}^c(S)} \lambda^{t_j} \langle \xi_\lambda, \xi_{\lambda'} \rangle \xi_{\lambda'} \right] \\ &= \sum_{\lambda \in \sigma_{\varepsilon,j}^c(S)} \lambda^{t_j} c_\lambda \xi_\lambda. \end{aligned}$$

Again, the range is closed by linearity and boundedness of $S^{\perp}_{\varepsilon,j}$ follows directly from boundedness of S.

(ii) Let $f \in V^{\varepsilon}_{j}$. We can write $f = \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} c_{\lambda}\xi_{\lambda}$. Then,

$$S^{\perp}_{\varepsilon,j} f \;\; = \;\; \sum_{\lambda \in \sigma_{\varepsilon,j}(S)} c_{\lambda} \left[\sum_{\lambda' \in \sigma^{c}_{\varepsilon,j}(S)} \lambda^{t_j} \left\langle \xi_{\lambda},\, \xi_{\lambda'} \right\rangle \xi_{\lambda'} \right].$$

Here again, we use orthonormality and the fact that $\sigma^{c}_{\varepsilon,j}(S) \subseteq \sigma_{\varepsilon,j}(S)$ to get

$$S_{\varepsilon,j} f = \sum_{\lambda' \in \sigma_{\varepsilon,j}(S)} \lambda^{t_j} \xi'_{\lambda}.$$

The arguments for statements for (iii) and (iv) are identical to Proposition 3.38. □

Remark 3.44. If $f \in V_0$, by construction, there exists a sequence $\{f_j\}_{j \in \mathcal{J}} \subseteq W^{\varepsilon}_{j}$ such that $f = \sum_{j \in \mathcal{J}} f_j$, and $\langle f_j,\, f_{j'} \rangle = 0$ if $j \neq j'$ [8].

Theorem 3.45. *Suppose that we have the frame multiresolution analysis from Theorem 3.39, and let W^{ε}_{j} be defined as in equation 3.29. Define*

$$\Psi := S^{\perp}_{\varepsilon,j} \Phi_j. \tag{3.32}$$

Then, $\bigcup_{j \geq 0} \Psi_j$ is a frame for V^{ε}_{0}.

Proof. We prove the theorem in two steps.

(i) First, we show that Ψ_j is a frame for W^{ε}_{j}. Let $f \in W^{\varepsilon}_{j}$. Then

$$\sum_{k \in \mathrm{K}} |\, \langle f,\, \psi_{j,k} \rangle \,|^2 = \sum_{k \in \mathrm{K}} |\, \left\langle f,\, S^{\perp}_{\varepsilon,j} S_{\varepsilon,j} \psi_{0,k} \right\rangle \,|^2 = \sum_{k \in \mathrm{K}} |\, \left\langle S_{\varepsilon,j} S^{\perp}_{\varepsilon,j} f,\, \psi_{0,k} \right\rangle \,|^2.$$

Now $S^{\perp}_{\varepsilon,j} f \in W^{\varepsilon}_{j} \in V^{\varepsilon}_{j}$. Hence, $S_{\varepsilon,j} S^{\perp}_{\varepsilon,j} f \in V^{\varepsilon}_{j} \subseteq V_0$. Thus, we have

$$C'_1 \| S_{\varepsilon,j} S^{\perp}_{\varepsilon,j} f \|^2 \leq \sum_{k \in \mathrm{K}} |\, \langle f,\, \psi_{j,k} \rangle \,|^2 \leq C'_2 \| S_{\varepsilon,j} S^{\perp}_{\varepsilon,j} f \|^2,$$

The upper bound is easily found as

$$\|S_{\varepsilon,j}S_{\varepsilon,j}^{\perp}f\|^2 \leq \|S_{\varepsilon,j}\|^2\|S_{\varepsilon,j}^{\perp}\|^2\|f\|^2 \leq \|f\|^2,$$

and for the lower bound, using the pseudoinverse, we obtain

$$\|S_{\varepsilon,j}S_{\varepsilon,j}^{\perp}f\| \geq \frac{\|f\|}{\|(S_{\varepsilon,j}^{\perp})^{\dagger}\|\|S_{\varepsilon,j}^{\dagger}\|}.$$

(ii) Next, we show that $\bigcup_{j\geq 0}\Psi_j$ is a frame for V_0^{ε}. Using Remark 3.44 gives us

$$\begin{aligned}\sum_{j\geq 0}\sum_{k\in \mathrm{K}}|\langle f,\,\psi_{j,k}\rangle|^2 &= \sum_{j\geq 0}\sum_{k\in \mathrm{K}}|\sum_{l\in\mathcal{L}}\langle f_l,\,\psi_{j,k}\rangle|^2 \\ &= \sum_{j\geq 0}\sum_{k\in \mathrm{K}}|\langle f_j,\,\psi_{j,k}\rangle|^2.\end{aligned}$$

By part (i), we have

$$C_1'\sum_{j\geq 0}\|f_j\|^2 \leq \sum_{j\geq 0}\sum_{k\in \mathrm{K}}|\langle f_j,\,\psi_{j,k}\rangle|^2 \leq C_2\sum_{j\geq 0}\|f_j\|^2.$$

But $\sum_j\|f_j\|^2 = \|f\|^2$, so our claim holds, i.e.,

$$C_1'\|f\|^2 \leq \sum_{j\geq 0}\sum_{k\in \mathrm{K}}|\langle f,\,\psi_{j,k}\rangle|^2 \leq C_2\|f\|^2.$$

□

3.4 Composite Diffusion Frames

As we mentioned before, diffusion wavelet frames gives us a higher degree of flexibility than diffusion wavelets. However, once a frame Φ_0 for the space V_0^{ε} is set, for a given symmetric diffusion semigroup $\{S^t\}_{t\geq 0}$ we have no more control on the functions $\Phi_j = S_{\varepsilon,j}\Phi_0$ than spans the approximation spaces V_j^{ε}. Depending on the properties of the functions that we are trying to represent, this could be non-efficient representation

system. This problem is not unlike trying to represent any function of $L^2(\mathbb{R})$ with a Haar system, or other wavelet systems lacking more or less properties that we are trying to capture in the function. In Chapter 2, we have studied composite dilations wavelet frames, which, depending on the composition of dilations, capture different properties of functions such as directionality. Our goal is to build upon that idea to construct diffusion systems with the ability of representing functions with different properties more efficiently.

Let us introduce composite diffusion frames.

Theorem 3.46. *(Composite diffusion frames MRA) Suppose that we have all the assumptions in Theorem 3.39. In addition, we have a compact symmetric diffusion semigroup $\{T^t\}_{t\geq 0}$. We assume that $\{S^t\}_{t\geq 0}$ and $\{T^t\}_{t\geq 0}$ are "similar" in the sense that S and T have identical spectrum and there exists an invertible U such that for each λ, the eigenfunctions ξ_λ and ζ_λ of, respectively, S and T satisfy*

$$\zeta_\lambda = U\xi_\lambda. \tag{3.33}$$

Then, for each $j \geq 0$, $t_j = 2^{t+1} - 1$, the sequence of frames

$$\Phi_j := \{S_{\varepsilon,j}T_{\varepsilon,i}\Phi_0\}_{i\in\mathcal{I}_j\subset\{0,1,\ldots j-1\}} \tag{3.34}$$

has the following properties:

(i) for each $j \geq 0$, $\langle\Phi_j\rangle = \left\langle V_j^\varepsilon\right\rangle$ and hence $\langle\Phi_{j+1}\rangle \subseteq \langle\Phi_j\rangle$,

(ii) for each $j \geq 0$, Φ_j is a frame for $\langle\Phi_j\rangle$ with frame constants $C_1''', C_2''' > 0$.

Proof. Let $S_{\varepsilon,j}$ be defined as in (3.26) and let

$$T_{\varepsilon,i}f = \sum_{\lambda\in\sigma_{\varepsilon,i}(T)} \lambda^{t_i} \langle f, \zeta_\lambda\rangle \zeta_\lambda \text{ for all } f \in L^2(X,\mu).$$

Note that all the properties of $S_{\varepsilon,j}$ holds for $T_{\varepsilon,i}$.

We prove our theorem in several steps.

a. For $j \geq 0$, $i \in \mathcal{I}_j$, $T_{\varepsilon,i}\Phi_0$ is a frame for

$$V_i^\varepsilon = \langle\{\xi_\lambda : \lambda \in \sigma_{\varepsilon,j}(S)(S)\}\rangle .$$

Define the subspace

$$Y_i^\varepsilon = \langle\{\zeta_\lambda : \lambda \in \sigma_{\varepsilon,j}(S)(S)\}\rangle .$$

We claim that by similarity of S and T, for all $i \geq$

$$V_i^\varepsilon = Y_i^\varepsilon. \tag{3.35}$$

This is because invertibility of U implies that the set $\{\xi_\lambda\}_{\lambda\in\sigma_{\varepsilon,i}(T)}$ $\{\zeta_\lambda\}_{\lambda\in\sigma_{\varepsilon,i}(T)}$ have the same span.

The first consequence of (3.35) if that, since Φ_0 is a frame for V_0^ε, then is it also a frame for Y_0^ε. Then, we can employ Theorem 3.39 to obtain $T_{\varepsilon,i}\Phi_0$ is a frame for Y_i^ε, and hence $T_{\varepsilon,i}\Phi_0$ is a frame for V_i^ε with frame bounds $C_1' = \dfrac{C_1}{\|T_{\varepsilon,j}^\dagger\|}$ and $C_2' = C_2$.

b. For $j \geq 0$, $i \in \mathcal{I}_j$, the composition $S_{\varepsilon,j}T_{\varepsilon,i}\Phi_0$ is a frame for V_j^ε. By part a., we have

$$\langle T_{\varepsilon,i}\Phi_0\rangle = Y_i^\varepsilon = V_i^\varepsilon.$$

So, via Proposition 3.38 (ii) and, again, the same argument as in Theorem 3.39,

we obtain that for each $i \in \mathcal{I}_j$, $S_{\varepsilon,j}T_{\varepsilon,i}\Phi_0$ is a frame for V_j^ε with frame constants $C_1'' = \dfrac{C_1'}{\|S_{\varepsilon,j}^\dagger\|}$ and $C_2'' = C_2'$.

c. Define Φ_j as in (3.34). Φ_j is therefore a finite union of frames (over the i's), which is a frame with frame constants $C1'''$, which is the minimum of C_1'' over $i's$ and $C_2''' = |\mathcal{I}_j||\mathcal{I}_j|C_2''$.

□

Here, we have in mind that U could be an operator such as a shear, or a unitary operation such as a rotation. Note that, by varying the powers $T_{\varepsilon,i}$, $i \in \mathcal{I}_j$, applied to our frame, we can select, to a certain extent, how we would like to modify the frame at each each j. We say to a certain extent, because, although we have gained in flexibility, U fixes all the types of transformations that we can make.

To have even more control at every level, consider the following theorem:

Corollary 3.47. *Suppose that we have all assumptions of Theorem 3.46 except that, instead of the compact diffusion semigroup T we are given a family of closed range, bounded operators $\{U_i\}_{i\in\mathcal{I}_j}$, $|\mathcal{I}_j| = N_j < \infty$, such that, for each $i \in \mathcal{I}_j$, $j' \le j$,*

$$\langle \{U_i\phi_k\}_{i\in\mathcal{I}_j,k\in\mathrm{K}}\rangle = \langle \{\xi_\lambda\}_{\lambda\in\sigma_{\varepsilon,j'}(S)}\rangle. \tag{3.36}$$

Then, for each $0 \le j < \infty$, $t_j = 2^{j+1}-1$, the sequence of frames $\Phi_j := \{S_{\varepsilon,j}U_i\Phi_0\}_{i\in\mathcal{I}_j}$, has the following properties:

(i) for each $j \ge 0$, $\langle\Phi_j\rangle = \langle V_j^\varepsilon\rangle$ and hence $\langle\Phi_{j+1}\rangle \subseteq \langle\Phi_j\rangle$,

(ii) for each $j \ge 0$, Φ_j is a frame for $\langle\Phi_j\rangle$ with frame constants $C_1', C_2' > 0$.

Proof. The argument for this proof is very similar to the previous. For $j = 0$, the result is again obvious. For the remaining j, note that for each $i \in \mathcal{I}_j$, condition 3.36 implies that $\{U_i\phi_{0,k}\}_{i\in\mathcal{I},k\in\mathrm{K}}$ a frame for V_i^ε, $i < j$ with frame bounds $\dfrac{C_1}{\min\limits_{i\in\mathcal{I}}\|U_i^\dagger\|}$ and $C_2|I_j|\max\limits_{i\in\mathcal{I}}\|U_i\|$. Then, we apply exactly the same reasoning as in the proof of Theorem 3.46 to obtain the results for V_j^ε. □

Based on Theorem 3.46 and Corollary 3.47, it would be easy to see how analogues of Theorem 3.41, approximations of V_j^ε, would be constructed for composite dilation diffusion frames MRA and also how we could construct diffusion frame MRA with composite operators. In addition, we can construct composite dilations wavelet frames in using an argument similar as before for a frame defined using composite dilations.

Remark 3.48. We would like to address the connection of Chapter 2 and Chapter 3. We have already noted that some common properties of S^t and D_A such as expansion of the support functions on f. Recall that in order to build the composite frames in Chapter 2, we used the assumption that $A \in \mathbf{G}_A$ normalizes $\mathbf{G}$ if, for each $B \in \mathbf{G}$, $ABA^{-1} \in \mathbf{G}$, where $\mathbf{G}$ be a finite subgroup of $SL_n(\mathbb{Z})$ and $\mathbf{G}_A = \{A^j : j \in \mathbb{Z}\} \subset GL_n(R)$. Here, the "normalization" condition becomes the "similarity" of operators S and T in the composite diffusion frames constructions. In both cases, we want the operators D_B and T^{t_j} to modify functions f_k in some V, which D_A^j and S^{t_j} will later dilate, in such a way that the span of f_k is left unchanged.

3.5 Examples

The obvious catch here is how to find the family $\{U_i\}_{i\in\mathcal{I}_j}$ that allows such a system to exist. Actually, this can be done by considering different $\{U_i\}_{i\in\mathcal{I}_j}$ such that Equation 3.33 holds for the eigenvectors of T and S for that t_j. We will show such a family of operators simply based on diffusion wavelets.

Example 3.49. *We can argue that the diffusion wavelets of Coifman and Maggioni are a particular case of composite operators diffusion wavelets. Indeed, they start with an orthonormal basis, which is just a special case of frame. And we can define the orthogonalization process as applying a sequence of operators $\{U_i\}_{i\in\mathcal{I}_j}$ to the orthonormal basis.*

Example 3.50. *(**Novel idea: Diffusion Shearlets**) Suppose that we have all assumptions of the composite diffusion frames theorem. Here, we are given an invertible, shear matrix U and form a family of operators T_k, similar to S, such that the eigenvectors of each T_k and the eigenvectors of S are related as follows:*

$$\zeta_{\lambda,k} = U^k \xi_\lambda.$$

The function $U^k\xi_\lambda$ are sheared versions of the ξ_λ. For $f \in L^2(X,\mu)$, we define

$$T_{\varepsilon,i,k} f = \sum_{\lambda\in\sigma_{\varepsilon,i}(T)} \lambda^{t_i} \langle f, \zeta_{\lambda,k}\rangle \zeta_{\lambda,k}.$$

Then, via Theorem 3.46, it is it easy to see that $\Phi_j = \{S_{\varepsilon,j} T_{\varepsilon,i,k}\Phi_0\}_{i\in\mathcal{I}_j\subset\{0,1,\ldots j-1\},k\in K_j\subset\mathbb{Z}}$ is a frame for V_j. We call the set of Φ_j diffusion shearlets.

Chapter 4

Detection of Anomaly In Human Retina using Laplacian Eigenmaps and Vectorized Matched Filtering

4.1 Introduction

Age-related Macular Degeneration (AMD) is the leading cause of blindness in elderly patients in industrialized nations [63]. The earliest clinically observable sign of retinal pigment epithelium dysfunction, the cause of AMD, is an accumulation of color fundus deposits known as drusen. Epidemiological studies have found that drusen increases in number and size with age and that larger, irregularly shaped, perifoveal drusen (soft drusen) signal a high risk for progression to advanced AMD. Currently, pathologists manually classify drusen based on size and shape, but there is a growing interest in automated, analytic tools, to diagnose AMD in early stages, track its progression and test the effectiveness of new treatment methods.

The classification of drusen can be complex because limitations of the imaging devices. These cause nonlinear mixing between spectral information about the chemicals of interest and other responses from other layers just beneath the retina, or from different illumination patterns [29]. In the work that has been done up to this

point, the trend is to detect macular drusen or other anomalies through direct fundus image segmentation or through application of a multiscale analysis procedure such as wavelet analysis or dimensionality reduction followed by a classification algorithm. In [70], the automatic segmentation of fundus images is done first by choosing a sample healthy area in the image using a region growing method, and then classifying blood vessels and other areas of the macula as healthy. For the final segmented image, the inverse image of the segmented image is generated as an unhealthy region of the macula. In [83], the authors propose an automated AMD detection system using discrete wavelet transform and feature ranking strategies based on, among others, the Kullback-Lieber Divergence and Bhattacharyya Distance. In [50], the authors classify drusen based on multiscale analysis and kernel-based drusen detection. They use the Mexican hat wavelet to obtain a feature vector composed of the response of the image to filtering and apply Support Vector Data Description to this feature vector to detect anomalies. In [88], a Gabor kernel-based filter bank is used to eliminate spurious regions which may be confused with drusen in fundus images. Each region is then represented with a number of features and classified as drusen and non-drusen using a hybrid classifier based on Naive Bayes and Support Vector Machine. Czaja, Ehler and al. use a great variety of techniques to respond to the challenges of drusen detection [4, 5, 45]. In [29], Czaja and Ehler propose Schrödinger Eigenmaps, a new semi-supervised manifold learning and recovery technique. Schrödinger Eigenmaps are a kernel method based on Schrödinger operators made of a Laplacian and a potential barrier constructed with carriers of labeled information. This method gives promising results on standard biomedical datasets and multispectral retinal images.

This extension of Laplacian eigenmaps is further studied in [58]. Ehler and al. also detect precursors in retinal images of AMD via such methods as quantification of retinal chromophores or sparse representation and variational methods [38, 44].

In this joined work with Wojciech Czaja, Lucia Simonelli and Denise Cunningham, we focus on detecting structures that represent any anomaly rather than drusen only, as the variability in color, shape or texture of drusen is very high among patients. We detect these anomalies in autofluorescence retinal images provided by the National Eye Institute at the NIH (National Institute of Health). Our generous collaborators at the NIH, Robert Bonner, Emily Chew and Denise Cunningham, also provided us with much insight on retinal anomaly detections via multiple conversations and through their work on AMD[21, 82, 95, 96]. Our algorithm focuses on distinguishing normal background features in the macula from drusen or other anomalies. This approach has been taken in previous work [50, 65, 87, 98].

In the detection algorithm that we propose, dimensionality reduction and feature enhancement is achieved by use of the Laplacian Eigenmaps (LE) of Belkin and Niyogi [2], and classification is done via vectorized matched filtering (VMF). The Laplacian Eigenmaps algorithm is an example of so called kernel-based techniques, which have many forms of utility [6, 24, 41, 58]. In [87], for example, Rajapakse applies them to a multitude of biomedical applications such as microarray gene expression data for nonlinear gene cluster analysis. Our motivation to use a spectral technique such as LE lies in the fact that given multispectral retinal imagery, LE has the capability to indirectly segment the multispectral information into coherence classes of pixels with similar spectral behavior. As such, it avoids the pitfalls of mixing

models that rely on presence of pure pixels, i.e., pixels that contain only one feature or class, in the image. The information obtained by applying the LE algorithm then takes the form of images formed by the eigenvectors of the Laplace operator defined on the data graph. By correlating spectral information, LE provides us with a way to select the most significant eigenvectors, i.e., the eigenvectors that emphasize anomaly structures. In other words, we exploit the capability of the LE algorithm to produce vectors (or images) that visually emphasize the locations of anomalies in retinal images.

We apply two matched filtering-based algorithms, Overlapping-detections Matched Filtering (OMF) and Vectorized Matched Filtering (VMF), to the Laplacian eigenmaps. The first, OMF is mainly used for comparison purposes. In OMF, filtering and detection are applied to each individual image, and then the anomaly detections are chosen as those that match in most images. In VMF, the filtering is applied to all eigenimages at once and the average response is used for a single detection. Our results demonstrate that VMF is superior in performance and hence, that operating on the retinal images or retinal eigenimages as data cubes is the better approach.

Our work is presented as follows. In Section 4.2, we give a review of Laplacian Eigenmaps and matched filtering. In Section 4.3, we describe our approach to anomaly detection using a Laplacian Eigenmaps algorithm (LE) followed by matched filtering-based algorithms (OMF and VMF). In Section 4.4, we outline the preparation of the images and explain the implementation of our method. In section 4.5, we present and discuss our results. Finally, in section 4.6, we summarize this work and give some ideas for future research.

4.2 Background

We begin with a review of Laplacian Eigenmaps, and make the connection with other kernel-based algorithms as well as diffusion semigroups. We follow with a review of matched filtering.

4.2.1 Laplacian Eigenmaps

Suppose that we are given the set of vectors $X = \{x_1, \cdots x_N\} \subset \mathbb{R}^D$, where D is large. We want to find vectors $Y = \{y_1, \ldots y_N\} \subset \mathbb{R}^d$, where $d << D$, such the important information in the original data $\mathcal{X}$ is retained in $\mathcal{Y}$. We assume that the data set X is sampled from a manifold $\mathcal{M}$ embedded in $\mathbb{R}^D$.

The steps of the LE algorithm are the following:

1. Adjacency matrix construction.

Given N vectors of dimension D, we construct an adjacency graph G that represents the data with nodes (or vertices) formed by the data vectors, and with edges that represent the distances between the nodes that are defined to be "close". There are two ways of establishing the notion of closeness for this neighborhood construction. In both cases, we start by computing the Euclidean distances between all pairs of points x_i, x_j, i, $j = 1, \ldots, N$. Then, one option is to define two nodes as connected by an edge if the Euclidean distances between them is less that some pre-defined, fixed ε. Although this option is geometrically intuitive, it can be difficult to choose an appropriate ε, and it often yields disconnected graphs. In our work, we use the second option: the **k-nearest neighbors method**. Here, the nodes are connected

if x_i is among the k nearest neighbors of x_j or vice versa. The main advantage of k nearest neighbors is that it is simpler to choose k and we have more control over the degree of connectivity of our graph. Also, just as with the first option, we have a symmetric relationship.

All the information for the graph G is stored in an adjacency matrix A.

2. Heat kernel as weights.

The adjacency matrix A is modified by assigning weights to the edges of the graphs. Let $\sigma > 0$ and weight the edges of the graphs using the **heat kernel** as follows

$$w_{i,j} = \begin{cases} e^{\frac{-\|x_i - x_j\|^2}{\sigma}} & \text{if } i \text{ and } j \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Now, note that by using this type of weights, we would need to determine the appropriate σ. Although they are simpler alternatives (for e.g., setting $w_{i,j} = 1$ if x_i and x_j are connected by an edge and $w_{i,j} = 0$ otherwise), which avoids choosing σ, the heat kernel is a better option from a geometrical point of view. Indeed, the Laplace operator that we define next, based on these weights, is analogous of the Laplace Beltrami operator on manifolds, whose eigenfunctions have properties desirable for embedding. For a complete discussion on the geometric implications choice as well as the relation between heat flow and the Laplace Beltrami operator on manifolds, refer to the original paper [2].

3. Minimization problem.

We define a diagonal matrix D such that $d_{i,i} = \sum_j w_{i,j}$ and consider the minimization

problem

$$\min_{y^T Dy=I} 1/2 \sum_{i,j} \|y_i - y_j\|^2 w_{i,j} = \min_{y^T Dy=I} \text{trace}\left(y^t L y\right), \tag{4.2}$$

where $L = D - W$: $d \times d$ Laplace operator and I: identity matrix. Assume that the graph G is connected. The Laplace operator is a symmetric, positive semidefinite matrix that can be thought of as an operator on functions defined on vertices of G. For $y^T Dy = I$, we solve the minimization problem by noticing that it is equivalent to finding the d minimal eigenvalues solutions of

$$L\xi = \lambda D\xi. \tag{4.3}$$

Hence, the Laplacian eigenmaps are the eigenvectors of the matrix L. We order the eigenvalues from lowest to highest, $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N-1}$, and let $\xi_0, \ldots \xi_{N-1}$ be the corresponding eigenvectors. We ignore the eigenvector ξ_0 corresponding to the eigenvalue $x_0 = 0$ and use the next d eigenvectors as coordinates for the d-dimensional embedding Euclidean space.

Laplacian eigenmaps in the context of Kernel Methods

In [59], the authors show that all kernel-based techniques, such as Diffusion Maps and Laplacian eigenmaps are just special cases of kernel PCA [91]. The following summary appeared in [26].

In kernel eigenmaps techniques, the goal is to find the solution to the optimiza-

tion problem

$$\min_f \{Q_1(f)\}, \tag{4.4}$$

$$\text{subject to } Q_2(f) = 1 \tag{4.5}$$

where $Q_1(f) = \sum_{x \in X} Q_x(f)$. Here, Q_2 and $\{Q_x, x \in X\}$ are symmetric positive-definite quadratic forms acting on functions f defined on X. The forms Q_x measure local variations of f and the form Q_2 are defined to normalize f. For example, in the Locally Linear Embedding (LLE) method [40], $Q_{\mathbf{x}}$ is the squared norm of difference between the vector $\mathbf{x}$ of points in X and the vector f which contains the weighted average of the neighbors of each x in $\mathbf{x}$. The functions f are local in the sense that for points lying outside a predefined neighborhood of each point, the weights are set to 0. The normalization Q_2 is given by the fact that the weights of all points in the neighborhood of x sum up to 1.

It is very easy to see that, setting $f = y$, Equation (4.2) gives $Q_1(f)$ and $Q_2(f) = 1$ is the same as $y^T D y = I$.

Laplacian Eigenmaps and Diffusion methods

In Chapter 3, we defined symmetric diffusion semigroups on a graph or manifold 3.14. Now, the Laplacian $L = D - W$ meets all diffusion semigroup properties except for contraction, so its powers do not define a symmetric diffusion semigroup. However, powers of the closely related normalized Laplacian, $D^{-1/2}(I - W)D^{-1/2}$ generate a symmetric diffusion semigroup in a discrete setting. In [87], it is shown that the Fourier eigenvectors in diffusion maps is just the same as the eigenvectors obtained in

Laplacian Eigenmaps, except for the fact that the former are weighted with respect to the eigenvalues of the transition matrix P associated with the Markov chain defined on the data set.

4.2.2 Matched Filtering

In signal processing, matched filtering is used to detect the presence of specific elements in an unknown signal using an appropriately designed linear filter. Matched filter are such that the transfer (or impulse response) function which processes a received signal minimizes the effect of noise [85]. Matched filtering can therefore be very useful in biomedical applications which are often characterized by low signal to noise ratio (SNR) [63]. Vinodh Rajapakse, Zachary Rom and Jonathan Franck have already tackled the detection of drusen using matched and we are grateful for their insight and base codes.

Let us define this problem in more technical terms. We consider the detection of an arbitrary 1-D real discrete-time system signal $s(t)$. Although the term signal usually has that connotation, we make the explicit assumption that $s(t)$ is finite-energy, i.e., $s(t) \in L^2(R)$. To ensure that $\hat{s}(t)$, the Fourier transform of s has a well-defined inverse, we make the further assumption that $L^2(R) \cap L^1(R)$. Now, let $s(t)$ have the form

$$x(t) = y(t - t_0) + n(t), \tag{4.6}$$

where for $t \in \mathbb{R}$, $y(t)$ is the signal (or pulse) that we want to detect, $n(t)$ is noise, and A and t_0 is an unknown propagation time delay constant. In general, the output of

a matched filter is a sharp peak in response to the presence of the desired signal (or pulse) at its input. Using this peak, we can find the time t_0, and amplitude A of the signal within the input. First, let's go through some terminology.

Consider an arbitrary signal $x(t)$ as defined in (4.6) and recall that the Fourier transform of a function $f(t) \in L^2(\mathbb{R})$ is given by

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-2\pi\omega t} \mathrm{d}\omega.$$

The filter output $x_0(t)$ when $x(t)$ passes through the filter $h(t)$ is given by

$$x_0(t) = x(t) \otimes h(t) := \int_R \hat{x}(\omega)\hat{h}(\omega) e^{j\omega t} \mathrm{d}\omega, \tag{4.7}$$

i.e., by the inverse Fourier transform of $\hat{x}(\omega)\hat{h}(\omega)$. By linearity, we immediately have

$$x_0(t) = y(t) \otimes h(t) + n(t) \otimes h(t).$$

We will denote the output signal component $y(t) \otimes h(t)$ as $y_0(t)$ and the output noise component $n(t) \otimes h(t)$ as $n_0(t)$. Then the peak output signal to noise ratio (SNR) is defined as

$$SNR = \frac{|y_0(t)|^2}{\overline{|n_0(t)|^2}}, \tag{4.8}$$

where $\overline{|n_0(t)|^2}$ is the average noise power.

Now, we can define the matched filter.

Definition 4.1. Given a finite energy signal $x(t)$ defined as in (4.6), a **matched filter** $h_{MF}(t)$ (or $\hat{h}_{MF}(\omega)$) is a linear filter that maximizes the peak SNR at its output $x_0(t)$ at some specified instant t_d relative to t_0.

We will omit the formal details of the derivation, which can be found in, e.g., [62], [51] and arrive directly at the conclusion: the SNR is maximized when

$$\hat{h}_{MF}(\omega) = k\frac{\hat{x}^*(\omega)}{\hat{r}_n(\omega)}e^{-j\omega t_d}, \tag{4.9}$$

where k is some arbitrary constant and $\hat{r}_n(\omega)$ is the Fourier transform of the auto-correlation function of the noise:

$$r_n(\tau) = \int_{\mathbb{R}} n(t)n^*(t-\tau)\mathrm{d}t.$$

If we assume that the noise is Gaussian, then $\hat{r}_n$ is constant. Then, $H_{MF}(\omega) = kX^*(\omega)e^{j\omega t_d}$ and using inverse transform, we have $h_{MF}(t) = kx^*(-t+t_d)$. Assume without loss of generality that $k = 1$, $t_d = 0$ (matched filter with zero-delay). Moreover, we are only interested in real valued signals, therefore we finally have

$$h_{MF}(t) = x(-t). \tag{4.10}$$

This equation means that the impulse response of our filter is exactly a reversed copy of the transmitted signal. In practice, a series of filters, which we call templates, are correlated with an unknown signal. Components of the templates yielding "high" correlation coefficients will be retained as components of the signal.

It is fairly obvious to extend the above concept to a higher dimensional, discrete case, and in particular to our 2-D case. Suppose that we form a template image T of size $p \times q$. We compute the normalized cross-correlation NCC of the matrices template and each of our images, yielding a matrix that contains correlation coefficients. For

each pixel $I(x, y)$ in the image I, the correlation coefficient is given by

$$NCC(x, y) = \frac{\sum_{i}^{p}\sum_{j}^{q} I(x+i, y+j)T(i, j)}{\sqrt{\sum_{i}^{p}\sum_{j}^{q}|I(x+i, y+j)|^2}\sqrt{\sum_{i}^{p}\sum_{j}^{q}|T(i, j)|^2}}. \tag{4.11}$$

We use the absolute values of this NCC matrix as the response matrix for each of our image. The methods that we describe in Section 4.3.2 will combine these filter responses from different images in two different ways.

4.3 Methodology

4.3.1 Feature Extraction using Laplacian Eigenmaps

The first part of our method consist in just applying the Laplacian eigenmaps algorithm as described in Section 4.2.1 to the set of images provided by the NIH.

Figure 4.1 provides an example in which the Laplacian Eigenmap algorithm is applied to a set of retinal images and illustrates how the eigenimages make the anomalies more obvious.

4.3.2 Classification using Matched-Filtering-based Algorithms

Consider that we have a number of images of the same eye and perform matched filtering for each of them as described in Section 4.2.2. We integrate the responses in two ways.

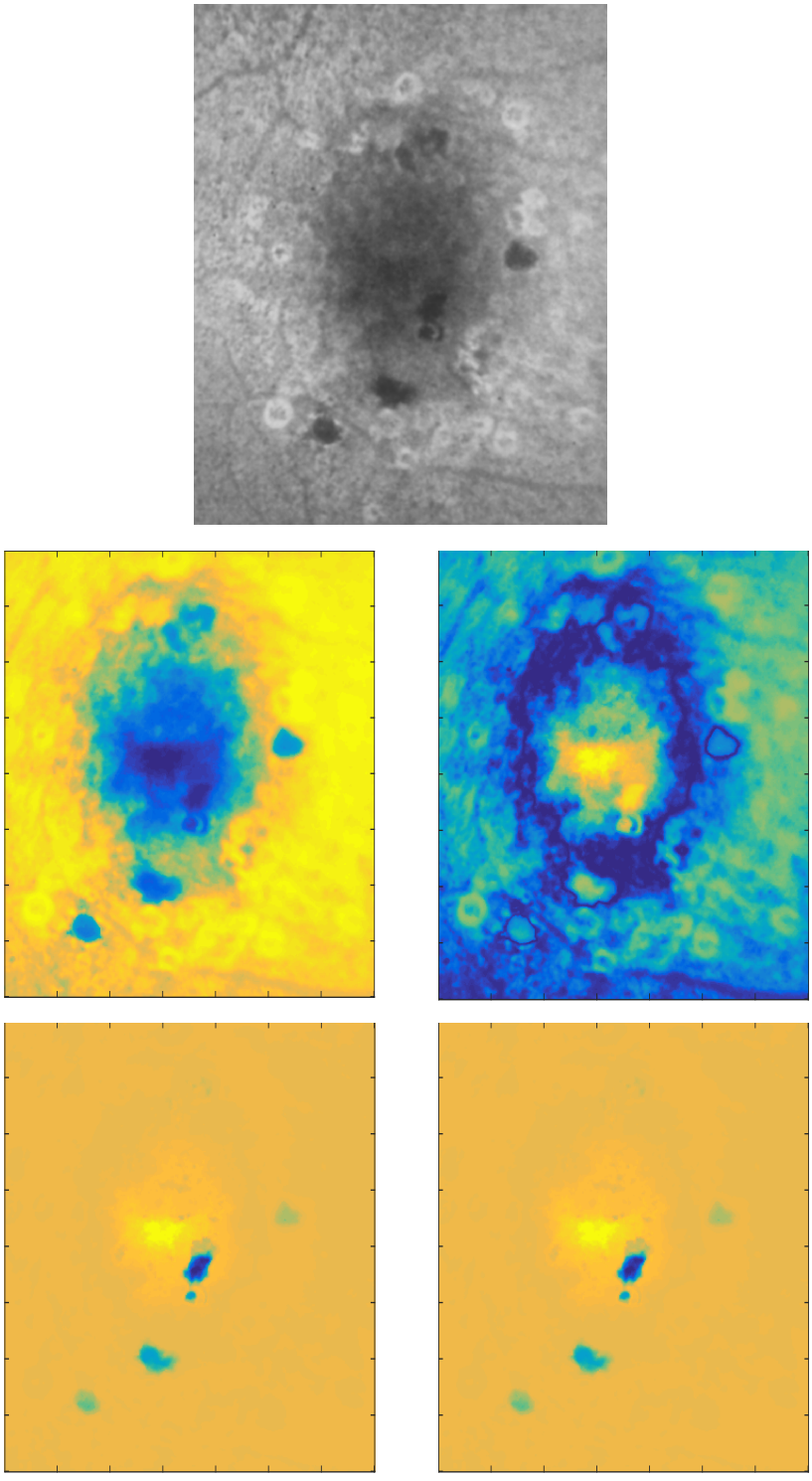


Figure 4.1: One of the original images in the green band and four eigenimages. Note that the eigenimages enhance the appearance of structures in the retinal images, pointing to the possibility of distinct anomaly spectral signatures. For example, the eigenimage on the second row, left, we can distinguish two classes of anomalies and the eigenimage the third row, right, emphasizes only structures with large dark centers surrounded by a thin contour with white fluffy appearance.

Overlapping-detections Matched Filtering (OMF)

We perform matched filtering for each image, identify detections for each individual image separately by applying a threshold to keep more significant correlations, and then keep the detections which overlap in an arbitrary number of images. The idea behind this method is that detections that persist across bands are less likely to be due to noise. The reason to keep detections present in a few images instead of all bands is that our experiments have shown that while the former gives reasonable results, the latter eliminates too many valid detections and is too sensitive to the fact that one of the bands might have images of poor quality that yield poor detections.

Vectorized Matched Filtering (VMF).

We perform matched filtering for each image, but this time, the final matched filter correlation matrix considered is obtained by taking the average of the correlation matrices across all images and applying a threshold to keep more significant correlations. The idea behind this method is, on one hand, that correct detections will lead to large correlations across all images and on the other hand, noisy detections will be more likely to produce small correlation coefficients in some images. Thus, the average response of the noisy pixels will be insignificant. In other work, the identification of anomaly in retina images was done via segmentation coupled with image enhancement approaches [70], [71]. Those classification techniques focus on single images, while with VMF, we are able to systematically incorporate all the information in all bands or eigenvectors at once.

4.4 Implementation

4.4.1 Image Preparation

One big part of our work is to turn the autofluorescence retinal images provided by the National Eye Institute (at the NIH) into data viable for the implementation. We begin with blue, green, and yellow AF images of the human retina. These colors are chosen for their ability to capture and absorb different chemicals. This was done in the following steps:

1. Registration

We register the 9 images of each eye (3 per color band, each of them with three RGB components) in order to determine the overlap between them and produce a common coordinate system. This is an extremely important step as misregistration can lead to translational, scale and rotational differences and hence, lower the performance of our algorithm.

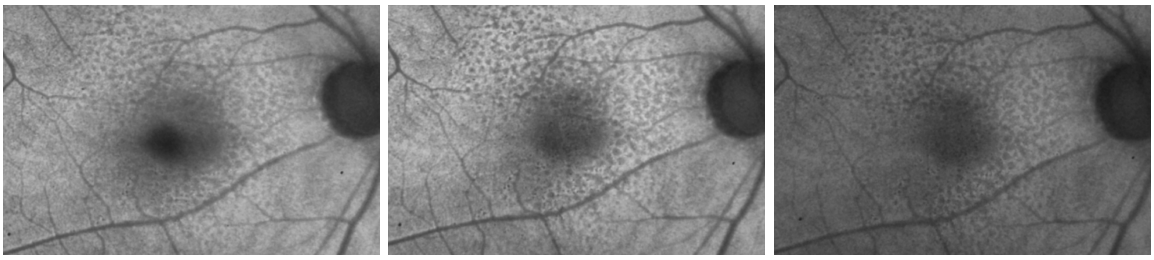
2. Alignment

After registration, we manually crop the images to a uniform size with an attempt to maintain consistency within a patient's eye image set. Following this, we align the images for each eye in order to generate a correspondence across the images. The images obtained after alignment have the same size and uniform overlap. The registration and alignment was done using the software i2k Align from DualAlign.

3. Vessel Mask

We create a mask of the image in order to remove contribution from the blood vessels, as they are otherwise selected by the detection algorithm as anomaly. Retinal vessel

extraction is in its own right a very active research area, and the results produced are quite impressive.



(a) One blue band component

(b) One green band component

(c) One yellow band component

(d) Vessel Mask

Figure 4.2: Blue, green, yellow autofluorescence images and corresponding vessel mask obtained using a simple vessel tracing technique with MATLAB commands `ginput` and `roipoly`.

In [39] for example, the authors develop a variational technique that approximates measured scanning laser ophthalmoscope image sets optimally within the range of the bleaching model. This technique is robust to noise and has consistent numerical results despite the differences in the variational settings. Other state of the art methods can be found in [92], [80] or [20], and [69] gives a useful survey of this topic. Here, we got the original idea to obtain our retinal vessel mask using a few simple steps in MATLAB. First, using the command `ginput` on a retinal image, we trace around the most visible vessels section by section. For each section, `ginput` returns the pixels

coordinates. We gather all the data points and use the command `roipoly` to create a polygon encasing these vessels. `roipoly` returns the binary image that we use as a mask for filtering. Figure 4.2 shows an example of one of the vessel masks created by using this method. Our method for creating vessel masks presents the advantage of being computationally efficient, avoiding complicated algorithms and the use of supplementary software. However, it is clearly not as precise as more sophisticated schemes. The most obvious weakness of our vessel extraction method can be seen in Figure 4.3, where most of incorrect detections lie on or near the small vessels that we did not capture in the vessel mask. As the purpose of this work is to focus on the role of nonlinearity and joint vector-valued analysis of retinal multispectral imagery, a simple (even if somewhat underperforming) method for vessel mask creation allows us to clearly separate and attribute the resulting artifacts.

```
                    Algorithm for vessel mask creation

// Input: grayscale retinal image IMG

// Output: binary colored vessel mask: MASK


% Display image

display(IMG) ⟵ imshow(IMG)

% Trace around points X, Y around vessels.

X, Y ⟵ ginput(cursor entries)

% Return region(s) of interest specified by the polygon described by points X, Y.

MASK ⟵ roipoly(IMG, X, Y)
```

4.4.2 Anomaly Detection

Our data set is made of the images obtained after registration et alignment of the retinal images of 3 patients. We picked these patients from a larger data set provided by the NIH based on the following criteria. First, we selected patients with a complete set of images in each band, ideally for both eyes. Second, their retinal images had to show significant levels of anomalies. Third and last, we wanted the anomalies to have different appearance, i.e., different shape and texture.

Laplacian Eigenmaps

For each patient, we crop the images to obtain smaller images of identical sizes $m \times n \times 3$ and then convert the images into grayscale $m \times n$ matrices. In the LE step, we vectorize these matrices to obtain $D = mn$ vectors in 9-dimensional space. We computed the LE using Laurens van der Maaten's Dimension Reduction Toolbox, which containing MATLAB implementations of several techniques for dimensionality reduction, some of which were optimized by Cloninger, Doster and Halevy. For the first patient, whom we will call Patient 1, the images were of size 325×350, and for the last two patients, Patient 2 and Patient 3, the images were of size 425×350. Hence, the set dataset is made of 113750 vectors in 9-dimensional space for Patient 1 and 148750 in 9-dimensional space for Patient 2. As we mentioned in the algorithm description, we choice of the parameters σ and k is key. For the parameter σ, we tried 10 values in several order of magnitude. Due the increase of computational cost as k increases, we tried 7 values between 5 and 100. The eigenimages shown in Figure 4.1 were obtained for $\sigma = 1$, $k = 30$.

Template creation

Our first algorithm step in the implementation step is the creation of templates. We rely on the following properties of drusen or anomalies in retinal images:

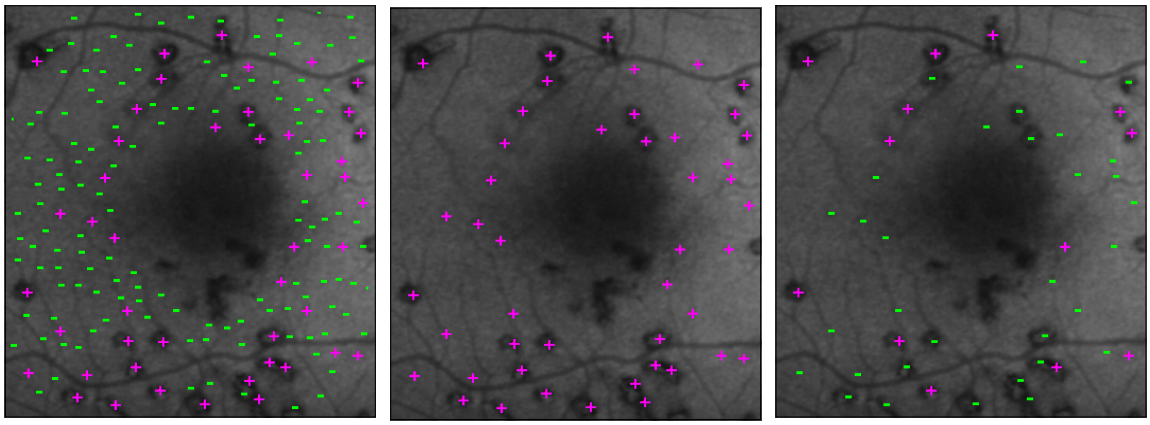
- The drusen or other anomalies usually have circular or ellipsoid shapes whose size does not vary significantly for an individual patient but varies significantly across patients.
- The centers of drusen or other anomalies have lower reflectance compared to other retinal surfaces, and therefore they appear darker relative to the background. However, this reflectance is comparable to that of the vessels or the fovea, which is the darkest part located in the centre of the macula. Therefore, we "remove" detections located on the vessels and fovea using the vessel masks.

Based on these observations, we produce a series of templates containing circles and ellipsoids of various sizes. We localize anomalies in the eye image of different patients, and then determine the approximate "radius," in pixels, of an anomaly. To account for different orientations, we rotate our templates by various angles between 0 and π. In our experiments, we determined that 12 rotation angles provide more than enough variations for the template orientation. Then, for each image, we compute the correlation with the different templates using the MATLAB routine `normxcorr` and retain the maximum response.

Thresholding and Detection

The term "detecting" anomalies precisely means determining their locations (or image coordinates). In our algorithm, given a maximum response matrix (correlations

retained after matched filtering), we first apply a threshold coefficient r to eliminate false detections due to some background elements that produce high correlation responses. Then, we get an intensity ranking by sorting responses by descending values. Following this ranking, we exclude responses that are in the neighborhood of a correlation "peak." At this stage, the radius of the neighborhood was defined by visual approximation of the anomaly size; we will propose an improvement of this method in the discussion. Note that the optimal threshold coefficient r^* is obtained by selecting a value that results in the most "reasonable" detections as determined by visual inspection. Figure 4.3 illustrates the process of determining the best r^* among the values $r = 0.50$, $r = 0.67$ and $r = 0.75$ for Patient 3. Here, one can see that $r = 0.67$ is the best since, on one hand, the extra detections (represented by the minus markers) obtained using $r = 0.50$ are mostly false detections, and on the other hand, using $r = 0.75$ excludes many valid detections. The optimal r is chosen as the highest value of r that strikes a good balance between maximizing the number of detections while minimizing the number of false detections. This optimal parameter depends on each patient and on the methods used for detection; in our experiments, we found that $0.60 \leq r^* \leq 0.70$.



(a) Detections at $r = 0.50$ versus 0.67: the minus markers are detected at $r = 0.50$ and not at $r = 0.67$

(b) Detections at 0.67.

(c) Detections at $r = 0.67$ versus 0.75: the minus markers are detected at $r = 0.67$ and not at $r = 0.75$

Figure 4.3: Variations of detections using different correlation thresholds.

To compare two methods in terms of matching or differing detections, or to determine if two detections overlap in OMF, we need to define a search neighborhood. Indeed, if we only match using identical pixel coordinates for the detection, we would not account for noise effect or slight variations in eye images that yield correlation peaks with slightly different coordinates which actually point to the same anomaly locations. So, for each correlation peak corresponding to a detection, in order to determine if detections overlap, the search neighborhood's radius for matching pixels is the same as the radius of exclusion from the neighborhood of a correlation "peak".

All the processing was performed using MATLAB R2014b on a Mac. OS X 10.10.

4.5 Results and Discussion

4.5.1 Detection in Absence of Added Noise

For all our results, any image containing only plus markers shows anomaly detections in this image. More precisely, these markers represent the anomalies' centers detected. In order to visually compare the effectiveness of different methods or different parameters, two images are created to track the anomaly centers detected by one method but not by the other. In these images, the plus markers in both images denote the anomalies detected by both methods while the minus markers denote anomalies detected by one method and not by the other.

First, we discuss the effect of the heat kernel and nearest neighbors parameters, respectively, σ and k on the quality of Laplacian eigenmaps. As noted in [2], there is not yet a principled way to choose the heat kernel. Although the authors attempt to demonstrate the relation between the choice of σ and the k, there is a lot that still needs to be done in that area. We will proceed mostly by trial and error. For low values of σ (0.1, 0.2), the Laplacian eigenmaps show very little structure. This is because the exponents in the heat kernels become large and all points are basically treated as far apart. Starting from $\sigma = 1$, we notice that the value of σ has very little effect on the images produced. The color scheme may change, but the structures present in the eigenimages do not. One way to improve this experiment could be to design an experiment to pick σ based on the distribution of the pixels. To pick k, we adopt the convention that the larger the number of pixels, the higher (although not at all proportionally) the number of nearest neighbors. For k small, there eigenimages have

little structure, but for larger k, the features are enhanced. Since the computational cost increases with k, we picked the smaller value in our experiments for which the eigenimages are almost similar to those at higher k.

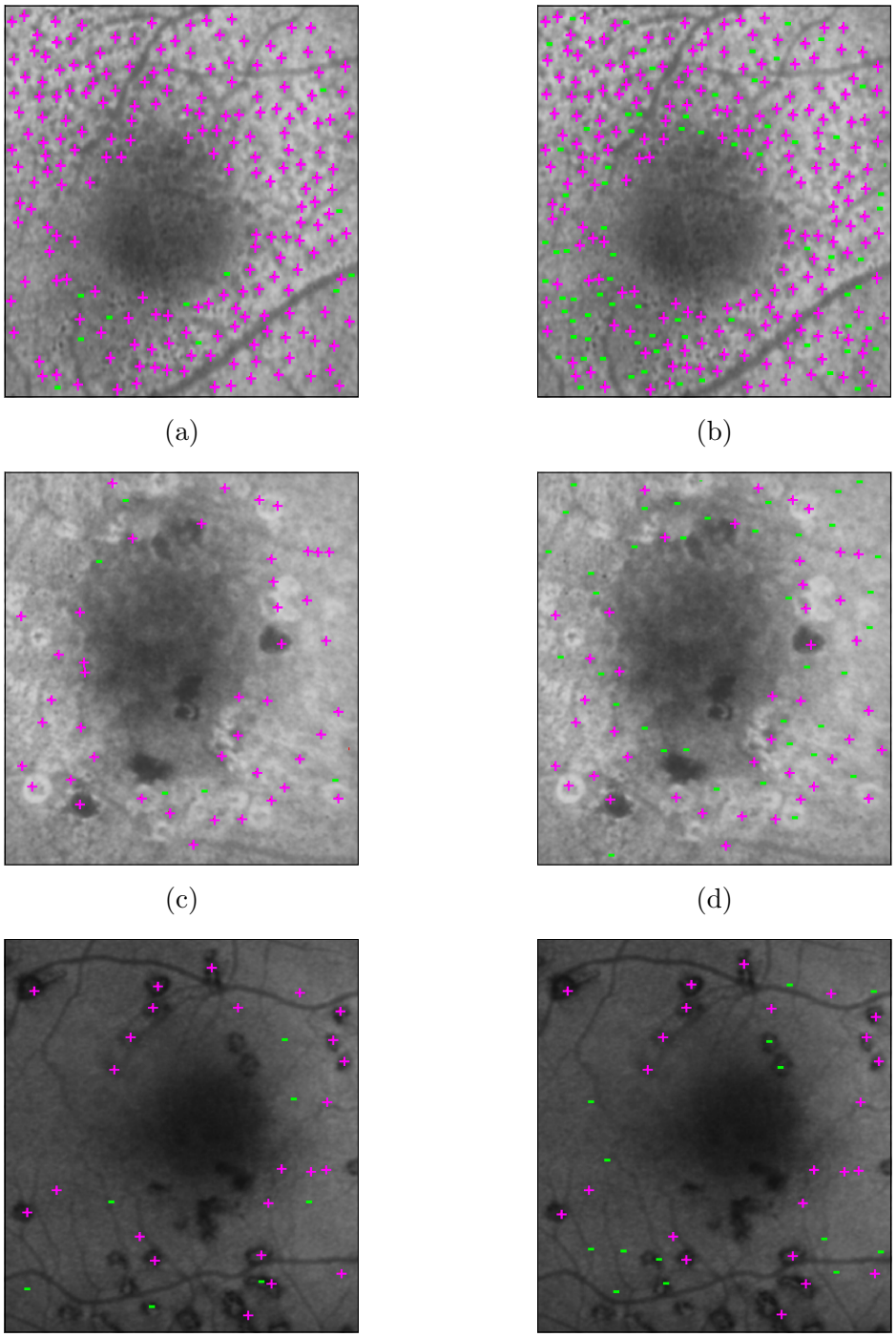
Next, we compare the matched filtering algorithms. We observe that VMF performs better than OMF when each is applied to the original images across all bands and then to the Laplacian Eigenmaps. The results are shown in Figures 4.4 and Figures 4.5, respectively, and summarized in Table 4.1. Given an optimal r^* for each method, Table 4.1 shows that VMF is able to maximize the number of correct detections and, when preceded by LE, yields a relatively low number of both false detections or missed detections, which we will henceforth refer to as false positives and true negatives, respectively. In particular, when OMF and VMF are preceded by LE (Figure 4.5), VMF clearly gives the best results with the highest number of correct detections and hence, lowest number of true negatives. The fact that OMF performs poorly when applied to Laplacian eigenimages is due to a mismatch in treating a joint data cube with a method that treat individual slices only. Each eigenimage in the joint data cube tends to highlight different structures and hence, there is little chance of overlapping detection. Now, observe that although the number of false positives is still significant, the ratio FP/C is the second lowest. Since most of the false positives occurred around vessels, we believe that with a more enhanced algorithm to create vessel masks, we will significantly lower this FP/C ratio. Table 4.1 gives detection results for only one patient, but these results are consistent with those obtained for other patients.

Table 4.1: Performances of OMF versus VMF

Type of Detections	OMF	VMF	LE-OMF	LE-VMF
Correct (C)	16	18	11	21
True Negative (TN)	6	4	11	1
False Positive (FP)	11	19	4	11
Rate of Correct anomaly Detection (R)	73%	82 %	50 %	95 %

Table 4.1 also shows the fact that the LE-VMF works better than VMF alone. As expected, the Laplacian eigenmaps are able to enhance the differences between anomalies and other parts of the retinal images. We also found that, when the eigenimages are chosen carefully, LE-VMF is less sensitive to changes in r. We picked an average of 4 eigenimages per patient. Note that, for each patient, the optimal r^* for LE-VMF is also the highest, hence, we have more certainty in the validity of the detections.

We can evaluate the performance of LE in preserving specific features of the data set by comparing its performance to one of the most highly acclaimed dimension reduction methods, Principal Component Analysis (PCA). Briefly, PCA can be used to simplify the analysis of high dimensional data sets (e.g. for data compression, visualization). It is a linear transformation that creates a new orthonormal basis for the data set. We will refer to the vectors of the PCA basis as the principal components. The PCA vectors are ordered such that the greatest variance of the data projection is contained in the lower order modes (first vectors) while the data variance residuals are contained in the higher order modes. In this way, PCA separates the data's important dynamics from redundant ones [61].



(a) (b) (c) (d) (e) (f)

Figure 4.4: Comparison of OMF and VMF performed on original images. For each patient, in both images the plus markers are the common detections. In the left images, the minuses mark the detections by OMF that were not detected by VMF and vice versa in the right images. VMF is able to pick up more anomalies, but both methods have a high rate of false positive.

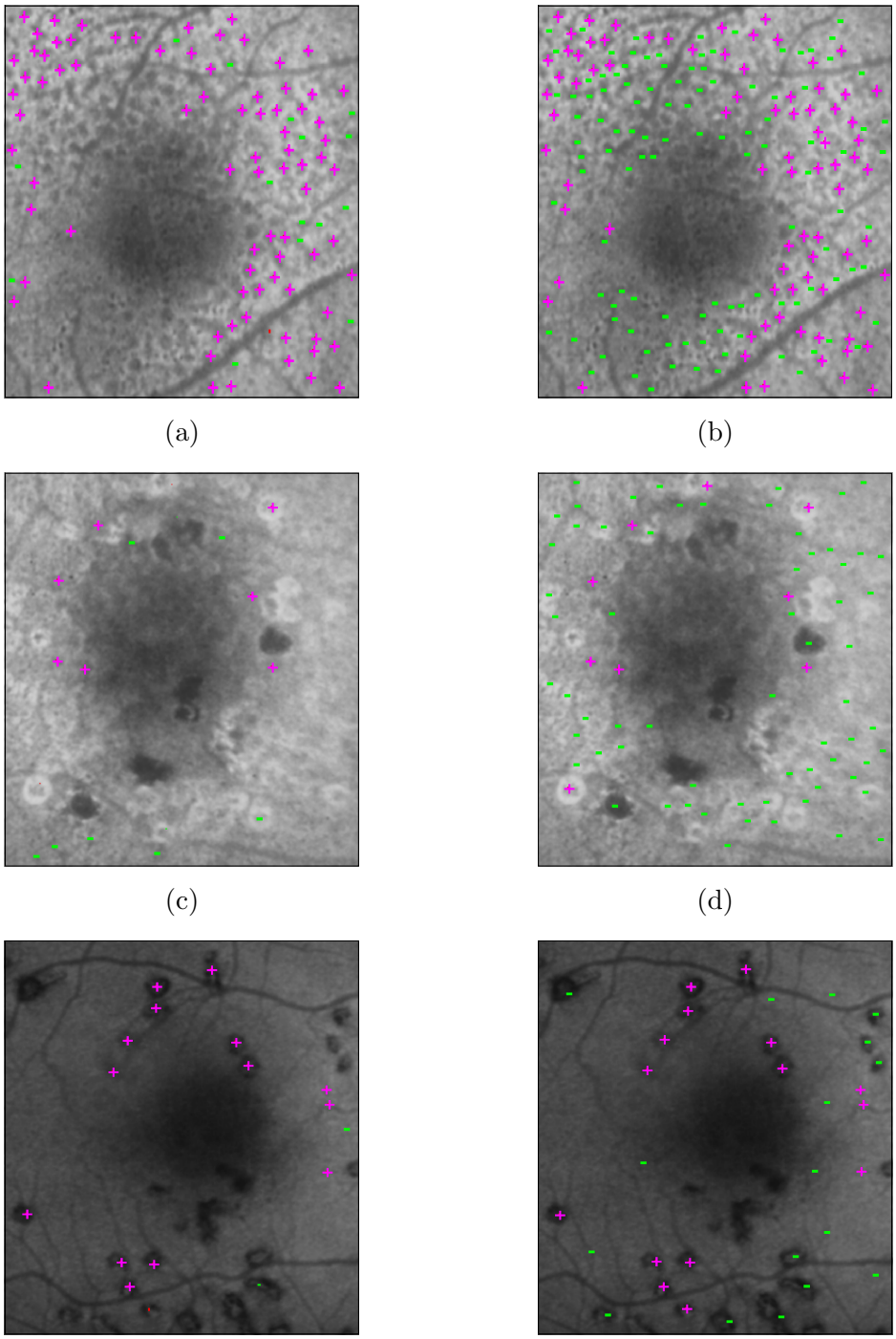


Figure 4.5: Comparison of OMF and VMF performed on Laplacian eigenimages. For each patient, in both images the plus markers are the common detections. In the left images, the minuses mark the detections by OMF that were not detected by VMF and vice versa in the right images. VMF performs a lot better than OMF and the rate of false positive is lower.

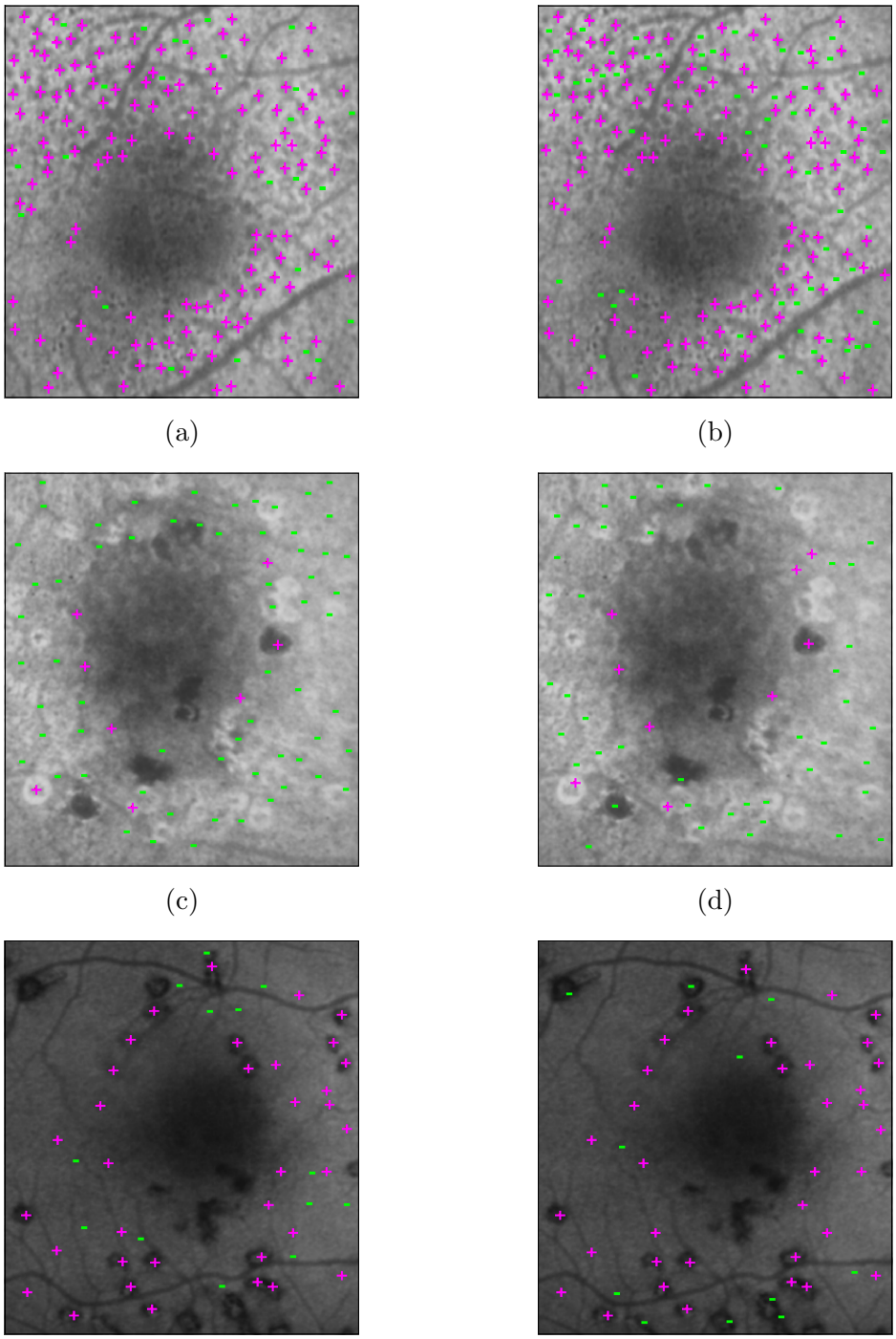


Figure 4.6: Comparison of PCA and LE as anomaly enhancing schemes. For each patient, in both images the plus markers are the common detections. In the left images, the minuses mark the detections by PCA-VMF that were not detected by LE-VMF and vice versa in the right images.

Table 4.2: Performances of VMF applied to principal components versus Laplacian eigenmaps.

Type of Detections	PCA-OMF	LE-VMF
C	16	21
TN	6	1
F	22	11
R	73 %	95 %

The principal components are found by computing the eigenvectors of the covariance matrix of the data zero-empirical mean. Figure 4.6 shows the performance of VMF on principal components versus VMF on Laplacian eigenmaps, and the detections for Patient 3 are summarized in Table 4.2.

For Patient 3, it is clear the LE-VMF performs better, but on other patients, for example, Patient 1, in which anomalies are less clear, LE-VMF seems to be hypersensitive and yields a lot of many false positives. However, this hypersensitivity becomes an advantage when different types of anomalies are mixed, for example, for Patient 2. In general, if we ignore false positives occurring in both cases because of our less than optimal vessels mask, the common detections of LE-PCA are correct. We can therefore think of a possible hybrid of LE and PCA for better detection.

4.5.2 Detection in Presence of Added Noise

Finally, we study how the presence of added (or artificial) noise can affect our results. To study the effect of noise on the performance of different algorithms for retinal images analysis, the standard choice is white Gaussian, [94], [76], [99]. We will therefore use white Gaussian noise w, with the added advantage that this is the assumption for

the type of noise in the matched filtering algorithm. Next, to decide which amount of noise to add to the images, we use the SNR. Since many signals have a very wide dynamic range, SNRs are often expressed using the logarithmic decibel scale. The SNR is measured in decibels as [78]:

$$SNR_{dB} = 10 \log_{10} SNR.$$

As we mentioned earlier, biomedical images are characterized by low SNR, therefore, we will use Gaussian noise such that we have relatively low SNR. As imaging procedure vary significantly, we could not find a typical (good) SNR for autofluorescence images. We therefore set the baseline at the approximate SNR value for our image. SNR estimation is in its own right a very active field (add reference). In this work, we simply use the native MATLAB function `snr` and found that the SNR of our image is about 30 decibels. So we evaluate our performance for SNR at 15dB, 20dB, 25dB, corresponding to noise variance of 2.5×10^{-4}, and 8×10^{-5}, 2.5×10^{-5}, respectively. Table 4.3 summarizes our findings. Overall, as can be expected, all the methods give less accurate detections in the presence of noise, but LE-VMF again outperforms all the other methods. Note that both methods involving classification in eigenimages instead of the original images perform better on all criterial. Another note is that PCA-VMF does best in terms of eliminating false positive, which reinforces our opinion that a hybrid method might work well. During the experiment, we observed that, in presence of noise, all the methods became a lot more sensitive to the choice of r^*. Since we analyzed only a few subjects, this was not a major issue, but in the future, when we will study much larger sets of images, we will need to

create a more "intelligent" way of determining this optimal r.

Table 4.3: Performances of different algorithms in presence of added noise.

SNR	Type of Detections	VMF	PCA-VMF	LE-VMF
25 dB	C	16	17	19
	TN	6	5	3
	FP	22	8	10
	R	73 %	77 %	86 %
20 dB	C	14	18	19
	TN	8	4	3
	FP	10	9	17
	R	64 %	68 %	86 %
15 dB	C	10	14	15
	TN	12	8	7
	FP	7	3	7
	R	45 %	64%	68%

In general, one area that would need improvement is eliminating false positives. We have a few ideas to do this. As we mentioned earlier, a number of far more sophisticated algorithms exist for more accurate retinal vessel extraction. Although these methods can be computationally inefficient, we expect that they could improve our final detection results significantly. Decreasing the number of false positive can could be also done by designing an automated thresholding scheme that would discriminate enhanced anomaly from other background elements. This would have the added advantage of further automating the detection process.

4.6 Conclusion

We have introduced a new method for automated detection of anomalies in autofluorescence retinal images. Our method takes advantages of the ability of LE to enhance the appearance of drusen or other anomalies in the images, before applying VMF. As a classification method, VMF treats the images as a data cube by averaging the matched filter responses over all the images. In this way, VMF gives more accurate results than a matched filtering algorithm such as OMF, which considers the common detections results from the filtering responses of individual images. Although LE is a nonlinear method and is relatively computationally expensive than say PCA, it is computed only once on the images and vastly outperforms PCA. Combined with VMF, which is linear and therefore, fast, our algorithm ends up with computational simplicity in addition to effectiveness in detection. Possible improvements could be attained by using a better vessel mask and templates made by a doctor-sanctioned approximation of real drusen shapes. For complete automation, in addition to a precise thresholding scheme, we could design a way of picking eigenimages with the most features without having to look at the images individually. We can also construct anomaly enhancements hybrid methods involving LE, PCA or experiment with other kernel-based methods. In particular, it would be interesting to study the performance of LE versus the Fourier basis and wavelets algorithms discussed in Chapter 3 and make a comparative analysis of these kernel-based methods.

A major goal of this work is also to establish a foundation for VMF analysis of Schrödinger Eigenmaps (SE), a semi-supervised version of LE. SE has been success-

fully applied to retinal multispectral data, see [29]. Our present method provides a groundwork for combining SE with vectorized matched filtering, in an effort to create a tool which will aid experts in the quantitative analysis of specific types of drusen such as hard versus soft or retinal versus reticular.

Bibliography

[1] E. Au-Yeung and J. J. Benedetto, *Balayage and short time fourier transform frames*, arXiv preprint arXiv:1309.0539, (2013).

[2] M. Belkin and P. Niyogi, *Laplacian Eigenmaps for dimensionality reduction and data representation*, Neural Computation, 15 (2002), pp. 1373–1396.

[3] J. J. Benedetto, *Harmonic analysis and applications*, vol. 23 of Studies in Advanced Mathematics, CRC Press, 1996.

[4] J. J. Benedetto, W. Czaja, and M. Ehler, *Frame potential classification algorithm for retinal data*, in 26th Southern Biomedical Engineering Conference, vol. 32 of IFMBE Proceedings, Springer Berlin Heidelberg, 2010, pp. 496–499.

[5] ——, *Wavelet packets for time-frequency analysis of multi-spectral images*, vol. 4, 2013, pp. 137–154.

[6] J. J. Benedetto, W. Czaja, J. Flake, and M. Hirn, *Frame based kernel methods for automatic classification in hyperspectral data.*, in IGARSS (4), 2009, pp. 697–700.

[7] J. J. Benedetto and S. Li, *Multiresolution analysis frames with applications*, in Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing: Digital Speech Processing - Volume III, ICASSP'93, Washington, DC, USA, 1993, IEEE Computer Society, pp. 304–307.

[8] ——, *The theory of multiresolution analysis frames and applications to filter banks*, Applied and Computational Harmonic Analysis, 5 (1998), pp. 389 – 427.

[9] J. J. Benedetto and J. Romero, *The construction of d-dimensional multiresolution analysis frames*, Journal of Applied Functional Analysis, 2 (2007).

[10] J. Blanchard, *Existence and accuracy results for composite dilation wavelets*, Washington University in St. Louis, 2007.

[11] J. Blanchard and K. Steffen, *Crystallographic Haar-type composite dilation wavelets*, in Wavelets and Multiscale Analysis, J. Cohen and A. I. Zayed, eds., Applied and Numerical Harmonic Analysis, Birkhuser Boston, 2011, pp. 83–108.

[12] J. Bremer, R. Coifman, M. Maggioni, and A. Szlam, *Biorthogonal diffusion wavelets for multiscale representations on manifolds and graphs*, in 59141M. SPIE, 2005, 2005.

[13] ——, *Diffusion wavelet packets*, Applied And Computational Harmonic Analysis, 21 (2006), pp. 95–112.

[14] H. Brézis, *Analyse fonctionnelle: théorie et applications*, Collection Mathématiques appliquées pour la maîtrise, Dunod, 1999.

[15] P. Burt and E. Adelson, *The Laplacian pyramid as a compact image code*, IEEE Transactions on Communications, 31 (1983), pp. 532–540.

[16] E. Candès, L. Demanet, D. Donoho, and L. Ying, *Fast discrete curvelet transforms*, Multiscale Modeling & Simulation, 5 (2006), pp. 861–899.

[17] P. Casazza, M. Fickus, J. Kovačević, M. Leon, and J. Tremain, *A physical interpretation of tight frames*, in Harmonic Analysis and Applications, C. Heil, ed., Applied and Numerical Harmonic Analysis, Birkhäuser Boston, 2006, pp. 51–76.

[18] P. Casazza and J. Kovačević, *Equal-norm tight frames with erasures*, Advances in Computational Mathematics, 18 (2003), pp. 387–430.

[19] P. Casazza, G. Kutyniok, and F. Philipp, *Introduction to finite frame theory*, in Finite Frames, P. G. Casazza and G. Kutyniok, eds., Applied and Numerical Harmonic Analysis, Birkhuser Boston, 2013.

[20] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, *Detection of blood vessels in retinal images using two-dimensional matched filters*, Medical Imaging, IEEE Transactions on, 8 (1989), pp. 263–269.

[21] E. Chew, T. Clemons, E. Agrón, R. Sperduto, J. SanGiovanni, M. Davis, and F. Ferris, *Ten-year follow-up of aAge-related Macular Degeneration in the age-related eye disease study: Areds report no. 36*, JAMA ophthalmology, 132 (2014), pp. 272–277.

[22] M. Christ, *Lectures on singular integral operators*, vol. 77 of CBMS Regional Conference Series in Mathematics, American Mathematical Society, Providence, RI, 1990.

[23] O. Christensen, *An introduction to frames and Riesz bases*, Applied and Numerical Harmonic Analysis, Birkhäuser Boston, 2003.

[24] A. Cloninger, *Exploiting Data-Dependent Structure for Improving Sensor Acquisition and Integration*, University of Maryland, College Park, 2014.

[25] R. Coifman, I. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, *Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems*, Multiscale Modeling & Simulation, 7 (2008), pp. 842–864.

[26] R. Coifman and S. Lafon, *Diffusion maps*, Applied and Computational Harmonic Analysis, 21 (2006), pp. 5–30.

[27] R. Coifman and M. Maggioni, *Diffusion wavelets*, Applied and Computational Harmonic Analysis, 21 (2006), pp. 53–94.

[28] R. Coifman, Y. Meyer, and V. Wickerhauser, *Adapted wave form analysis, wavelet-packets and applications*, in Proceedings of the second international conference on Industrial and applied mathematics, ICIAM 91, Philadelphia, PA, USA, 1992, Society for Industrial and Applied Mathematics, pp. 41–50.

[29] W. Czaja and M. Ehler, *Schrödinger Eigenmaps for the analysis of biomedical data*, CoRR, abs/1102.4086 (2011).

[30] W. Czaja and E. King, *Anisotropic shearlet transforms for $L^2(R^k)$*, Mathematische Nachrichten, 287 (2014), pp. 903–916.

[31] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Communications on Pure and Applied Mathematics, 41 (1988), pp. 909–996.

[32] ——, *The wavelet transform, time-frequency localization and signal analysis*, Information Theory, IEEE Transactions on, 36 (1990), pp. 961–1005.

[33] ——, *Ten lectures on wavelets*, no. 61 in CBMS/NSF Series in Applied Math., SIAM, 1992.

[34] I. Daubechies, A. Grossmann, and Y. Meyer, *Painless nonorthogonal expansions*, Journal of Mathematical Physics, 27 (1986), pp. 1271–1283.

[35] E. B. Davies, *Heat Kernels and Spectral Theory*, Cambridge University Press, 1989. Cambridge Books Online.

[36] D. Deng and Y. Han, *Harmonic Analysis on Spaces of Homogeneous Type*, Springer-Verlag Berlin Heidelberg, 2009. Springer-Verlag Berlin Heidelberg.

[37] M. Do and M. Vetterli, *Contourlets: a directional multiresolution image representation.*, in ICIP (1), 2002, pp. 357–360.

[38] J. Dobrosotskaya, M. Ehler, E. King, R. Bonner, and W. Czaja, *Sparse representation and variational methods in retinal image processing*, in 26th Southern Biomedical Engineering Conference, vol. 32 of IFMBE Proceedings, Springer Berlin Heidelberg, 2010, pp. 361–364.

[39] J. Dobrosotskaya, M. Ehler, E. J. King, R. F. Bonner, and W. Czaja, *Modeling of the rhodopsin bleaching with variational analysis of retinal images*, in Medical Imaging 2011: Image Processing, May 2011.

[40] D. Donoho and C. Grimes, *Hessian Eigenmaps: New locally linear embedding techniques for high-dimensional data*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 5591–5596.

[41] T. Doster, *Harmonic Analysis Inspired Data Fusion For Applications in Remote Sensing*, University of Maryland, College Park, 2014.

[42] R. Duffin and A. Schaeffer, *A class of nonharmonic Fourier series*, Transactions of the American Mathematical Society, 72 (1952), pp. 341+.

[43] G. Easley, D. Labate, and F. Colonna, *Shearlet-based total variation diffusion for denoising*, Image Processing, IEEE Transactions on, 18 (2009), pp. 260–268.

[44] M. Ehler, J. Dobrosotskaya, E. King, and R. Bonner, *Quantification of retinal chromophores through autofluorescence imaging to identify precursors of age-related macular degeneration*, in Excursions in Harmonic Analysis, Volume 2, Applied and Numerical Harmonic Analysis, Birkhuser Boston, 2013, pp. 355–371.

[45] M. Ehler, Z. Majumdar, E. King, J. Dobrosotskaya, E. Chew, W. Wong, D. Cunningham, W. Czaja, and R. Bonner, *High-resolution autofluorescence imaging for mapping molecular processes within the human retina*, in 26th Southern Biomedical Engineering Conference, vol. 32 of IFMBE Proceedings, Springer Berlin Heidelberg, 2010, pp. 344–347.

[46] K.-J. Engel and R. Nagel, *A Short Course on Operator Semigroups*, Universitext, Springer New York, 2006.

[47] L. Evans, *Partial differential equations*, no. 61 in Graduate Studies in Mathematics, American Mathematical Society, Providence, Rhode Island, 2000.

[48] G. Folland, *Fourier analysis and its applications*, Pure and applied undergraduate texts, American Mathematical Society, 1992.

[49] J. Fourier, *Théorie analytique de la chaleur; translated by Alexander Freeman; edited for the Syndics of the University Press.*, Cambridge [Eng.] :University Press,, 1822. http://www.biodiversitylibrary.org/bibliography/18544.

[50] D. Freund, N. Bressler, and P. Burlina, *Automated detection of drusen in the macula*, in Biomedical Imaging: From Nano to Macro, 2009 IEEE International Symposium on, June 2009, pp. 61–64.

[51] D. Friedman, *Detection of Signals by Template Matching*, Johns Hopkins Press, 1969.

[52] K. Friedrichs, *The identity of weak and strong extensions of differential operators*, Transactions of the American Mathematical Society, 55 (1944), pp. 132–151.

[53] K. Gröchenig, *Foundations of time-frequency analysis*, Applied and Numerical Harmonic Analysis, Birkhäuser Boston, 2001.

[54] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer, *Alpha molecules: curvelets, shearlets, ridgelets, and beyond*, vol. 8858, 2013, pp. 885804–885804–12.

[55] K. Guo and D. Labate, *Optimally sparse multidimensional representation using shearlets*, SIAM Journal on Mathematical Analysis, 39 (2007), pp. 298–318.

[56] K. Guo, D. Labate, W.-Q. Lim, G. Weiss, and E. Wilson, *Wavelets with composite dilations and their MRA properties*, Applied and Computational Harmonic Analysis, 20 (2006), pp. 202 – 236.

[57] A. Haar, *Zur theorie der orthogonalen funktionen systeme*, Mathematische Annalen, 69 (1910), pp. 331–371.

[58] A. Halevy, *Extensions of Laplacian Eigenmaps for manifold learning*, University of Maryland, College Park, 2011.

[59] J. Ham, D. Lee, S. Mika, and B. Schölkopf, *A kernel view of the dimensionality reduction of manifolds*, in Proceedings of the Twenty-first International Conference on Machine Learning, ICML '04, New York, NY, USA, 2004, ACM, pp. 47–55.

[60] D. Han, K. Kornelson, D. Larson, and E. Weber, *Frames for Undergraduates*, Student mathematical library, American Mathematical Society, 2007.

[61] H. Harman, *Modern Factor Analysis*, University of Chicago Press, 1976.

[62] S. Haykin, *Communication Systems*, Wiley Publishing, 5th ed., 2009.

[63] F. Holz, S. Schmitz-Valckenberg, R. Spaide, and A. E. Bird, *Atlas of Fundus Autofluorescence Imaging*, Graduate studies in mathematics, Springer Berlin Heidelberg, 2007.

[64] P. Jones, M. Maggioni, and R. Schul, *Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels*, Proceedings of the National Academy of Sciences, 105 (2008), pp. 1803–1808.

[65] S. Kankanahalli, P. Burlina, Y. Wolfson, D. Freund, and N. Bressler, *Automated classification of severity of age-related macular degeneration from fundus photographs*, Investigative ophthalmology & visual science, 54 (2013), pp. 1789–1796.

[66] Y. Katznelson, *An introduction to harmonic analysis*, Cambridge University Press, 1976.

[67] E. King, *Wavelet and frame theory: frame bounds gaps, generalized shearlets, grassmannian fusion frames and p-adic wavelets*, University of Maryland, College Park, 2009.

[68] E. King, G. Kutyniok, and X. Zhuang, *Analysis of inpainting via clustered sparsity and microlocal analysis*, Journal of mathematical imaging and vision, 48 (2014), pp. 205–234.

[69] C. Kirbas and F. Quek, *A review of vessel extraction techniques and algorithms*, ACM Comput. Surv., 36 (2004), pp. 81–121.

[70] C. Kose, W. S. Evik, and O. Gencalioglu, *Automatic segmentation of Age-related Macular Degeneration in retinal fundus images*, Computers in Biology and Medicine, 38 (2008), p. 611619.

[71] ——, *A statistical segmentation method for measuring age related macular degeneration in retinal fundus images*, Journal of Medical Systems, 34 (2008), pp. 1–13.

[72] A. Krishtal, D. Benjamin, L. Guido, and N. Edward, *Some simple Haar-type wavelets in higher dimensions*, The Journal of Geometric Analysis, 17 (2007).

[73] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, *Sparse multidimensional representation using shearlets*, vol. 5914, 2005, pp. 59140U–59140U–9.

[74] J. Lee and M. Verleysen, *Nonlinear dimensionality reduction*, Springer Publishing Company, Incorporated, 1st ed., 2007.

[75] P. Li and S. T. Yau, *On the parabolic kernel of the Schrödinger operator*, Acta Mathematica, 156 (1986), pp. 153–201.

[76] Z. Lu and B. Dosher, *External noise distinguishes attention mechanisms*, Vision Research, 38 (1998), pp. 1183 – 1198.

[77] S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 674–693.

[78] ——, *A wavelet tour of signal processing*, AP Professional, London, 1997.

[79] B. Manning, *Composite Multiresolution Analysis wavelets*, Washington University in St. Louis, 2012.

[80] A. Mendonca and A. Campilho, *Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction*, Medical Imaging, IEEE Transactions on, 25 (2006), pp. 1200–1213.

[81] Y. Meyer and R. Ryan, *Wavelets: Algorithms and Applications*, Society for Industrial and Applied Mathematics, 1993.

[82] S. Meyers, M. Ostrovsky, and R. Bonner, *A model of spectral filtering to reduce photochemical damage in age-related macular degeneration*, Transactions of the American Ophthalmological Society, 102 (2004), p. 83.

[83] M. Mookiah, R. Acharya, J. Koh, C. Chua, J. Tan, V. Chandran, C. Lim, K. Noronha, A. Laude, and L. Tong, *Decision support system for Age-related Macular Degeneration using discrete wavelet transform*, Medical & biological engineering & computing, 52 (2014), pp. 781–796.

[84] A. Nahmod, *Geometry of operators and spectral analysis*, Yale University, 1991.

[85] D. North, *An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems*, Proceedings of the IEEE, 51 (1963), pp. 1016–1027.

[86] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences, Springer New York, 1992.

[87] V. Rajapakse, *Data Representation for Learning and Information Fusion in Bioinformatics*, University of Maryland, College Park, 2013.

[88] G. Raza, M. Rafique, A. Tariq, and M. Akram, *Hybrid classifier based drusen detection in colored fundus images*, in Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference, Dec 2013, pp. 1–5.

[89] S. Roweis and L. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science, 290 (2000), pp. 2323–2326.

[90] H. Royden, *Real Analysis*, Mathematics and statistics, Macmillan, 1988.

[91] B. Schlkopf, A. Smola, J. Alexander, and K. Müller, *Kernel principal component analysis*, Advances in kernel methods: support vector learning, (1999), pp. 327–352.

[92] J. Soares, J. Leandro, R. Cesar, H. Jelinek, and M. Cree, *Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification*, Medical Imaging, IEEE Transactions on, 25 (2006), pp. 1214–1222.

[93] E. M. Stein and P. University, *Topics in Harmonic Analysis, Related to the Littlewood-Paley Theory*, Annals of mathematics studies, Princeton University Press, 1970.

[94] Y. Tolias and S. Panas, *A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering*, Medical Imaging, IEEE Transactions on, 17 (1998), pp. 263–273.

[95] B. Toy, N. Krishnadev, M. Indaram, D. Cunningham, C. Cukras, E. Chew, and W. Wong, *Drusen regression is associated with local changes in fundus autofluorescence in intermediate Age-related Macular Degeneration*, American Journal of Ophthalmology, 156 (2013), pp. 532 – 542.e1.

[96] Z. Wu, C. Luu, L. Ayton, J. Goh, L. Lucci, W. Hubbard, J. Hageman, G. Hageman, and R. Guymer, *Fundus autofluorescence characteristics of nascent geographic atrophy in Age-related Macular Degeneration*, Investigative Ophthalmology and Visual Science, 56 (2015), pp. 1546–1552.

[97] K. Yacoubou Djima, L. Simonelli, D. Cunnigham, and W. Czaja, *Detection of anomaly in human retina using Laplacian Eigenmaps and vectorized matched filtering*, vol. 9413, Mar 2015.

[98] Z. Yuanjie, W. Hongzhi, W. Jue, G. Jianbin, and J. Gee, *Multiscale analysis revisited: Detection of drusen and vessel in digital retinal images*, in Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, March 2011, pp. 689–692.

[99] F. Zana and J.-C. Klein, *Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation*, Image Processing, IEEE Transactions on, 10 (2001), pp. 1010–1019.