ABSTRACT

Title of dissertation: A STUDY OF THE RELATIONSHIP
BETWEEN SPECTRUM AND GEOMETRY
THROUGH FOURIER FRAMES AND
LAPLACIAN EIGENMAPS

Kevin W. Duke, Doctor of Philosophy, 2012

Dissertation directed by: Professor John J. Benedetto
Department of Mathematics

This thesis has two parts. The first is a study of Fourier frames. We follow the development of the theory, beginning with its classical roots in non-uniform sampling in Paley-Wiener Spaces, to its current state, the study of the spectral properties of finite measures on locally compact abelian groups. The aim of our study is to classify measures by their spectral properties. To this end, we extend a law of pure type and the classification of compactly supported, absolutely continuous spectral measures to the setting of locally compact abelian groups, using the generalization of Beurling density established by Gröchenig, Kutyniok, and Seip.

We also aim to understand the relationship between the geometry of the supporting set of a measure and the spectral properties it exhibits. To this end, we propose a method of constructing Bessel spectra for the middle-third Cantor measure.

In the second part, we study extensions of the Laplacian Eigenmaps algorithm and their uses in hyperspectral image analysis. In particular, we show that there is a

natural way of including spatial information in the analysis that improves classification results. We also provide evidence supporting the use of Schrödinger Eigenmaps as a semisupervised tool for feature extraction. Finally, we show that Schrödinger Eigenmaps provides a platform for fusing Laplacian Eigenmaps with other clustering techniques, such as kmeans clustering.

A STUDY OF THE RELATIONSHIP BETWEEN SPECTRUM AND GEOMETRY THROUGH FOURIER FRAMES AND LAPLACIAN EIGENMAPS

by

Kevin W. Duke

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
2012

Advisory Committee:
Professor John J. Benedetto, Chair/Advisor
Professor Alexander Barg
Professor Radu Balan
Professor Joel M. Cohen
Professor Matei Machedon

Dedication

In memory of Charlie, Lois, Dolly, and Bill.

Acknowledgments

Meaningless! Meaningless! says the Teacher. Utterly meaningless! Everything is meaningless. What do people gain from all their labors at which they toil under the sun? Generations come and generations go, but the earth remains forever.

-Ecclesiastes 1:1-4

Father God, thank you for giving my life meaning. Through your love, you have taught me about faith, humility, and community. The greatest achievement a man can aspire to is not something he toils over, etched in stone or erected from steel. Rather, it is something shared freely with those around him. The opportunity to be with those around us is a gift, each and every day.

The measure of a mathematician is not a discrete quantity (no, not even Dirac). In that respect, I would like to thank all those who have been with me these past six years.

First, my wife, who has loved me unconditionally—bearded, sleep-deprived, grungy, grumpy, pudgy, you loved me, and I love you. My family, the Dukes and the Strehles, for all of your love and support. The youth at Cherrydale UMC, for keeping me grounded and giving me something to look forward to every Sunday night.

To the Norbert Wiener Center community, especially John, Wojtek, and Kasso, without your generous support, I would not have made it this far. It has been an honor to work with and represent such a great group of people. John, thank you

for all of the time you have spent with me around the table, at the blackboard, or just passing by the lab.

And finally, speaking of the lab, thanks to all who made the long hours and terrible chairs more enjoyable: Flake, Avner, Nate, Enrico, Chae, Jens, Kevin H., Rongrong, Tim, Travis, Julia, and Weiwei, I wish you all the best.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Through analogy and generalization, the term *spectrum* has evolved into its own spectrum of mathematical meaning. Ambiguous words have their benefits, however, as they provide enough wiggle room to unite distinct ideas under a common theme. Case in point, the thesis in hand.

We use *spectrum* in the manner that Physics intended, as a set of observable or measurable quantities that provide clues as to the structure and composition of a system. In both parts, a spectrum will define a sequence of functions in the Hilbert space of square integrable functions with respect to some measure. We will study the relationship between such spectra and the geometry of the underlying space on which the functions are defined.

1.1.1 Fourier Analysis on Locally Compact Abelian Groups

Part I takes place on locally compact abelian groups (LCAGs). This is a classical area of mathematics and hence there are a number of excellent treatises on the subject. We used the following as resources for the necessary background material [25, 49, 50].

Definition 1.1.1 (Dual Group)**.** Let G be a LCAG. Continuous homomorphisms

mapping G to $\mathbb{C}$ of modulus 1 are called *characters*. The set of characters form a group, Γ, called the *dual group* of G. For each $\gamma \in \Gamma$, we write the character it is associated with as $e_\gamma(x)$. When $G = \mathbb{R}^d$, Γ is isomorphic to $\mathbb{R}^d$, and we will use the following normalization for the characters:

$$e_\gamma(x) = e^{2\pi i \langle \gamma, x \rangle}.$$

Endowed with the weak topology induced by G, Γ is a LCAG. The Pontryagin duality theorem tells us that duality is symmetric, that is, Γ is the dual of G if and only if G is the dual of Γ.

Definition 1.1.2 (Haar Measure)**.** Each G admits a family of nontrivial, translation invariant, Borel regular measures. Such measures are called *Haar measures* and are unique up to a multiplicative constant. We will denote Haar measures on G by m_G and Haar measures on Γ by μ_Γ. When $G = \mathbb{R}^d$, the Lebesgue measure is a Haar measure and we will denote it by m.

We will occasionally use the following notation for the integration of Borel measurable functions f on G and ϕ on Γ:

$$m_G(f) = \int_G f(x) dx, \ \mu_\Gamma(\phi) = \int_\Gamma \phi(\gamma) d\gamma.$$

Definition 1.1.3 (Fourier Transform)**.** The *Fourier transform* of a function $f \in L^1(G)$ is given by

$$\widehat{f}(\gamma) = \int_G f(x) \overline{e_\gamma}(x) dx.$$

Given a Haar measure on either G or Γ, a dual Haar measure can be chosen so that the Inversion Theorem holds.

Theorem 1.1.1 (Inversion Theorem)**.** *If* $f \in L^1(G)$ *and* $\widehat{f} \in L^1(\Gamma)$,

$$f(x) = \int_\Gamma \widehat{f}(\gamma)e_x(\gamma)d\gamma,$$

for almost every $x \in G$.

When the Inversion Theorem is satisfied, it can be used to show that there is a nice extension of the Fourier transform from $L^1 \cap L^2(G)$ to all of $L^2(G)$.

Theorem 1.1.2 (Plancherel's Theorem)**.** *There is a unique extension of the Fourier transform from* $L^1 \cap L^2(G)$ *to an isometry mapping* $L^2(G)$ *onto* $L^2(\Gamma)$.

We will use the same notation, $\widehat{f}$, for the Fourier transform of a function $f \in L^2(G)$.

1.1.2   Frames and Sampling

Frames, introduced by Duffin and Schaeffer in [12], provide a natural generalization of orthonormal bases in Hilbert spaces. Christensen's book [9] provides a good overview of basic frame theory and Young's book [54] is an excellent introduction to Fourier frames and related topics.

Frames provide a generalization of orthonormal bases by relaxing Parseval's identity:

$$\sum |\langle x, h \rangle|^2 = \|x\|^2.$$

Definition 1.1.4 (Frames)**.** Let $\mathbb{H}$ be a Hilbert space (finite or infinite dimensional). A set $F = \{f_j : j \in J\}$ is a *frame* for $\mathbb{H}$ if $\exists A, B > 0$ such that $\forall h \in \mathbb{H}$,

$$A\|h\|^2 \leq \sum_{j \in J} |\langle h, f_j \rangle|^2 \leq B\|h\|^2. \tag{1.1}$$

The tightest bounds satisfying (1.1) are called the *lower and upper frame bounds* for F. The upper bound is also called the *Bessel bound*. A set F is called a *Bessel sequence* if the Bessel bound is satisfied.

While frames are computationally more challenging than orthonormal bases, the relaxation of Parseval's identity has the advantage of allowing redundancy. As a result, frames provide a degree of robustness and numerical stability, necessary for many applications, that orthonormal bases do not have.

Definition 1.1.5 (Frame Operators). Associated with each frame is the *frame operator*

$$\mathcal{F} : \; \mathbb{H} \to \mathbb{H}, \quad \mathcal{F}(h) = \sum_{j \in J} \langle h, f_j \rangle f_j \tag{1.2}$$

The frame operator is bounded, self-adjoint, positive and invertible. Moreover, $\mathcal{F}^{-1}(F)$ is also a frame, called the *dual frame*, with frame bounds B^{-1} and A^{-1}. Each element of the Hilbert space can be expanded in terms of the frame and the dual frame.

Theorem 1.1.3 (Frame Expansion). *Let $h \in \mathbb{H}$. Then we have*

$$h = \sum_{j \in J} \langle h, \mathcal{F}^{-1} f_j \rangle f_j = \sum_{j \in J} \langle h, f_j \rangle \mathcal{F}^{-1} f_j. \tag{1.3}$$

In general, the coefficients appearing in (1.3) are not unique. When the expansion is unique, i.e. when the frame is a basis, we say that F is a *Riesz basis*.

Definition 1.1.6 (Fourier Frames). Let $M_1(\Gamma)$ be the the space of Borel probability measures defined on Γ. For $\mu \in M_1(\Gamma)$, we consider $\mathbb{H} = L^2(\mu)$, the space of square

integrable functions with respect to μ. When μ is the normalized restriction of Haar measure to a set Ω of finite measure, we will simply write $L^2(\Omega)$.

A *Fourier frame* is a sequence of characters which form a frame for $L^2(\mu)$. We will typically write such sequences as

$$E(S) = \{e_s(\gamma): \ s \in S \subset G\}.$$

If $E(S)$ is a

- Bessel sequence, we say that S is a *Bessel spectrum.*

- Fourier frame, we say that S is a *frame spectrum* and μ is a *spectral measure.*

- Riesz basis, we say that S is a *Riesz spectrum* and μ is a *Riesz spectral measure.*

- Fourier basis, i.e. orthonormal basis, we say that S is a *Fourier spectrum* and μ is a *Fourier measure.*

If μ is Haar measure restricted to some set Ω, we say Ω has these properties and replace *measure* in the terms above with *set.*

Our definition of a spectral measure differs from that found in the literature. Typically, the term spectral measure is reserved for a measure that simply admits a Fourier basis. We are interested in studying Fourier frames in general, however, and we found it more convenient to use the term to refer to this broader class of measures.

Fourier frames arise naturally in sampling theory for spaces of bandlimited functions.

Definition 1.1.7 (Paley-Wiener Spaces)**.** Let $f \in L^2(G)$. We say that f is Ω-bandlimited if $\mathrm{supp}(\widehat{f}) \subset \Omega$;

$$PW_\Omega = \{f \in L^2(\Omega) : \ \mathrm{supp}(\widehat{f}) \subset \Omega\}$$

is the space of Ω-*bandlimited functions*. These spaces are also often referred to as *Paley-Wiener spaces* because of the work done by Paley and Wiener in the setting where Ω is an interval in $\mathbb{R}$ [46].

The Classical Sampling Theorem states that if a function is bandlimited to an interval, then it is completely determined by its sampled values, taken uniformly at a rate greater than or equal to its bandwidth, i.e. the length of the interval. Fourier frames provide a generalization of the Classical Sampling Theorem to non-uniform sampling.

1.1.3 Iterated Function Systems

Definition 1.1.8 (Iterated Function System)**.** An *iterated function system* (IFS)

$$\Phi = \{\phi_1, \ldots, \phi_N\}$$

is a collection of contraction mappings on a complete metric space X.

Definition 1.1.9 (IFS Operators)**.** An iterated function system induces a mapping on the power set of X, $\mathcal{P}(X)$, given by

$$Y \in P(X) \mapsto \Phi(Y) = \bigcup_{i=1}^{N} Y_i$$

$$Y_i = \phi_i(Y). \tag{1.4}$$

This mapping is called the *Hutchinson operator.* A set is said to be *invariant* with respect to Φ if

$$Y = \Phi(Y) = \bigcup_{i=1}^{N} Y_i.$$

Iterated function systems also induce a mapping on $M(X)$. For $\mu \in M(X)$, the *push forward* of μ by $\phi_i \in \Phi$ is given by

$$\phi_i(\mu)(Y) = \mu(\phi_i^{-1}(Y)). \qquad (1.5)$$

Let $c \in \mathbb{R}^N$ be a *probability vector*; that is, for $1 \le i \le N$, $c_i > 0$ and

$$\sum_{i=1}^{N} c_i = 1.$$

For each Φ and c, let $\Phi_c : M_1(X) \to M_1(X)$ be given by

$$\Phi_c(\mu) = \sum_{i=1}^{N} c_i \mu_i.$$

A measure is said to be *invariant* with respect to Φ_c if

$$\mu = \Phi_c(\mu) = \sum_{i=1}^{N} c_i \mu_i.$$

The interest in iterated function systems lies in the study of their invariants. In [26], Hutchinson showed that each IFS has a unique compact invariant set and a unique invariant measure. In particular, iterated function systems can be used to create measures supported by sets with a fractal-like structure.

An IFS of interest to us will be

$$\Phi = \{\frac{1}{3}x, \ \frac{1}{3}(x+2)\};$$

Φ produces the *middle-third Cantor set* as an invariant. The invariant measure supported by the middle-third Cantor set is called the *middle-third Cantor measure.*

7

1.1.4   Laplacian Eigenmaps

Part II of the thesis focuses on the application of Laplacian Eigenmaps to the problem of material classification in hyperspectral images.

Laplacian Eigenmaps (LE) was proposed as a tool for dimension reduction and feature extraction in [1]. It is motivated largely by the theoretical connections between the eigenvalues of the Laplacian and the geometric properties of the system it is acting on provided by spectral graph theory and spectral geometry.

Definition 1.1.10 (Graph Laplacian)**.** Given a graph, let G be the *adjacency matrix* for the graph. That is, $G(i, j) = 1$ if there is an edge connecting the i-th and j-th nodes, and $G(i, j) = 0$ otherwise. A *weight matrix* for G is an assignment of positive weights to the edges of G. That is, $W(i, j) > 0$ if $G(i, j) = 1$ and is 0 otherwise. The *degree matrix*, D, for W is the diagonal matrix storing the total amount of weight present at each node:

$$D(i) = \sum_j W(i, j).$$

The *weighted graph Laplacian* is given by

$$L = D - W.$$

Definition 1.1.11 (LE Algorithm)**.** The algorithm for Laplacian Eigenmaps consists of three steps:

1. Construct a nearest neighbor graph for the given dataset.

2. Compute a weighted graph Laplacian.

3. Solve a generalized eigenvalue problem.

1.2 Outline of Results

In Section 2.1, we prove an oversampled version of the Classical Sampling Theorem in the setting of locally compact abelian groups. When $G = \mathbb{R}^d$, we show that oversampling allows smooth kernels to be used in the sampling reconstruction formula. In Section 2.2, we introduce the theory of non-uniform sampling in LCAGs. In particular, we make note of the extension of Landau's density criterion for sampling [37] to compactly generated LCAGs [22]. We prove that the classical stability theorems for sampling in PW_Ω, $\Omega \subset \mathbb{R}^d$, can be extended to sampling in compactly generated LCAGs.

In Section 3.1, we introduce the main objectives of our general research program:

- Characterize measures of each spectral type.

- Given $\mu \in M_1(\Gamma)$, provide a complete characterization of its spectra.

Using the results from Section 2.2, we extend to compactly generated LCAGs the recent work in [24], which shows that $\mu \in M_1(\mathbb{R}^d)$ is spectral only if it is discrete, absolutely continuous, or singularly continuous. We also extend to compactly generated LCAGs the complete characterization, provided in [36], of compactly supported, absolutely continuous spectral measures.

In Section 3.2, we provide a simple patch to an apparent gap in Hutchinson's proof for invariant measures, noted, for example, in [34]. We show that, without loss of generality, it can be assumed we are working in a situation where his proof is correct. We also introduce Cantor measures and show that three different constructions

of Cantor measures:

- as invariants of iterated function systems,

- as the distributional derivative of the Cantor-Lebesgue function on a perfect homogeneous set, and

- as the Haar measure on a totally disconnected group,

overlap nontrivially. In particular, they provide three equivalent constructions of the middle-third Cantor measure.

In Section 3.3, we examine the spectral properties of the middle-third Cantor measure. Known to lie outside the class of Fourier measures, it is a significant open problem as to whether the middle-third Cantor measure admits a Fourier frame. We present a simple, in terms of the theory involved, viable approach towards constructing Bessel spectra for the middle-third Cantor measure. The idea is a natural one and was motivated by the construction of the Cantor set itself. Hence, it is unsurprising that it is not unique. The work in this section was done independently, however, it overlaps with a recent paper of Dutkay, Weber, and Han [13].

In Section 4.1, we provide a brief overview of the theory for Laplacian Eigenmaps. In Section 4.2, we present two examples of its use in image analysis. The first provides the details of a research project, born out of a collaboration with the Naval Surface Warfare Center, Panama City, in which we used Laplacian Eigenmaps in the analysis of sonar data in an effort to determine the contents of a barrel on the ocean floor. The second is an example of its use in the analysis of hyperspectral images.

The latter example provides a basis of comparison for the methods introduced in Chapter 5.

In Section 5.1, we describe a natural way in which spatial information can be included in the analysis of hyperspectral images via Laplacian Eigenmaps, significantly improving the accuracy of the classification of materials in the image. In Section 5.2, we introduce Schroedinger Eigenmaps (SE), a generalization of Laplacian Eigenmaps, presented by Czaja and Ehler in [11]. We provide a proof of concept that SE can be used as a semi-supervised learning tool for feature extraction in hyperspectral images. We also show that Schroedinger Eigenmaps provides an interesting platform in which Laplacian Eigenmaps can be fused with clustering techniques, such as Kmeans Clustering. This fusion of methods yields a significant improvement in classification accuracy on a dataset where Laplacian Eigenmaps performs poorly.

Part I

Fourier Frames

Chapter 2

Fourier Sampling

2.1 Uniform Sampling

The cornerstone of sampling theory is the Classical Sampling Theorem. It states that, given $\Omega = [-\omega, \omega]$, every function $f \in PW_\Omega$ can be recovered from its sampled values if they are taken at a rate greater than or equal to the bandwidth, 2ω. The minimal rate for recovery, determined by the bandwidth, is called the *Nyquist rate.*

Theorem 2.1.1 (Classical Sampling Theorem). *Let $\frac{1}{T} \geq 2\omega$, then*

$$\forall f \in PW_\Omega, \ f(x) = T \sum f(nT) \frac{\sin(2\omega\pi(x - nT))}{\pi(x - nT)} \qquad (2.1)$$

where the series converges in $L^2(\mathbb{R})$ and uniformly on $\mathbb{R}$.

The Classical Sampling Theorem is often referred to as the Shannon Sampling Theorem, the Nyquist-Shannon Sampling Theorem, the Whittaker-Shannon-Kotelnikov Sampling Theorem, or some combination thereof. The ideas contained in the theorem actually date back at least as far as Cauchy [5].

When sampling at the Nyquist rate, there is only one possible sampling kernel that can be used in the reconstruction, namely the sinc function given in (2.1). In [4], Benedetto shows that by relaxing the Nyquist condition, a broad family of sampling kernels can be used. In this section, we prove a version of this theorem

for LCAGs. The benefits of oversampling are easily evident in $\mathbb{R}^d$, where smooth sampling kernels can be chosen to replace sinc in the reconstruction formula. We provide a simple construction of a family of smooth sampling kernels.

2.1.1 Oversampling in Locally Compact Abelian Groups

Kluvánek generalized the Classical Sampling Theorem, proving it in the setting of locally compact abelian groups [32]. To sample uniformly in a LCAG, values are taken over a lattice.

Definition 2.1.1 (Lattices). A *lattice* is a discrete subgroup of G whose quotient group, G/H, is compact. When $G = \mathbb{R}^d$, this is equivalent to $H = A\mathbb{Z}^d$, where $A \in GL_d(\mathbb{R})$. Such an A is called a *basis* for H. Lattice bases are not unique; given one basis A for H, every basis can be written as AV, where $V \in GL_d(\mathbb{Z})$.

Definition 2.1.2 (Tiles). Given a lattice H, let C_0 be any measurable coset representation of G/H. We call such a C_0 a *tile* for H, since we can write G as the pairwise disjoint union of translates of C_0 by the members of H. When $G = \mathbb{R}^d$, the sets $C_0 = AQ_0$, where A is a basis for H and $Q_0 = [0,1]^d$ is the unit cube, are tiles for H. These tiles are called the *fundamental cells* for H.

Not all LCAGs contain a lattice. For example, $\mathbb{Q}_p$, the field of p-adic numbers, fails to contain a discrete subgroup. The following structure theorem for compactly generated, LCAGs shows that assuming either G or Γ is compactly generated is sufficient for G to contain a lattice. See [25], e.g., for a proof.

Theorem 2.1.2 (Structure Theorem)**.** *Let G be a compactly generated, locally compact abelian group. Then G is isomorphic to $\mathbb{R}^d \times \mathbb{Z}^n \times K$, where $n, d \geq 0$ and K is a compact abelian group.*

Definition 2.1.3 (Dual Lattice)**.** The *annihilator* of a closed subgroup H is defined as

$$H^{\perp} = \{\gamma \in \Gamma : e_{\gamma}\big|_H \equiv 1\}. \tag{2.2}$$

To keep our notation consistent, we will use Λ as a pseudonym for $H^{\perp}$. The annihilator Λ is the dual group of G/H. Similarly, H is the dual group of Γ/Λ. A locally compact abelian group is discrete if and only if its dual is compact. It follows that H is a lattice if and only if Λ is also a lattice. Thus, the annihilator of a lattice is often called the *dual lattice.*

A key concept in Kluvánek's generalization is periodization by the dual lattice.

Definition 2.1.4 (Λ-Periodization)**.** Given a lattice H with dual lattice Λ and a function ϕ defined on Γ, the Λ-*periodization of ϕ* is the function

$$\phi_{\Lambda}([\gamma]) = \sum_{\Lambda} \phi(\lambda + \gamma) \tag{2.3}$$

defined on Γ/Λ.

Let $\mu_{\Gamma/\Lambda}$ be the Haar measure on Γ/Λ of unit mass. The mapping $\phi \mapsto \mu_{\Gamma/\Lambda}(\phi)$ is a translation invariant, positive linear functional on $C_c(\Gamma)$. Hence, there is a constant $c_H > 0$ such that

$$c_H \int_{\Gamma} \phi(\gamma) d\gamma = \int_{\Gamma/\Lambda} \phi_{\Lambda}([\gamma]) d[\gamma]. \tag{2.4}$$

In fact, $c_H = (m_\Gamma(\Omega))^{-1}$, where Ω is any tile for Λ.

We can now state Kluvánek's theorem.

Theorem 2.1.3 (Kluvánek, [32]). *Let $H \subset G$ be a lattice with dual lattice Λ. If $\Omega \subset \Gamma$ is a tile for Λ, then*

$$\forall f \in PW_\Omega, \ f(x) = c_H \sum_{h \in H} f(h) k_\Omega(x - h), \qquad (2.5)$$

where k_Ω is defined by $\widehat{k_\Omega} = \mathbb{1}_\Omega$. The series converges in $L^2(G)$ and uniformly on G.

The Nyquist rate in this setting corresponds to $1/c_H$. However, whereas the Nyquist rate uniquely defines the sampling set in the Classical Sampling Theorem, in general, there can be many lattices with the same Nyquist rate. What makes the Kluvánek Theorem (and thus the Classical Sampling Theorem) work is not the sampling rate itself, but the relationship between Γ/Λ and its dual H.

By relaxing this relationship slightly, we can adapt Kluvánek's argument to prove the following.

Theorem 2.1.4. *Let Ω_0 be a relatively compact, measurable subset of Γ. Let $H \subset G$ be a lattice with dual lattice Λ such that the canonical mapping $\pi : \Gamma \to \Gamma/\Lambda$ is injective on Ω_0. Let $k \in L^2(G)$ be such that*

(i) $\widehat{k}\big|_{\Omega_0} \equiv 1$, and

(ii) π is injective almost everywhere on $\mathrm{supp}(\widehat{k})$.

Then

$$\forall f \in PW_{\Omega_0}, \ f(x) = c_H \sum f(h) k(x - h). \qquad (2.6)$$

16

The reconstruction converges in $L^2(G)$ and uniformly on G.

Proof. Let Ω be a tile for Λ whose closure contains $\mathrm{supp}(\widehat{k})$. We will think of $L^2(\Omega)$ and similar spaces as closed subspaces of $L^2(\Gamma)$, with each $\phi \in L^2(\Omega)$ vanishing off of Ω.

Renormalize μ_Γ, letting $\widetilde{\mu}_\Gamma = c_H \mu_\Gamma$. It follows from (2.3) that Λ-periodization provides an isometric isomorphism between $L^2(\Omega)$ and $L^2(\Gamma/\Lambda)$.

Let $\phi_h(\gamma) = \mathbb{1}_\Omega(\gamma)e_h(\gamma)$. The Λ-periodization of the set $\{\phi_h : h \in H\}$ is precisely the group of characters on Γ/Λ. Since Γ/Λ is compact, and since we have chosen $\mu_{\Gamma/\Lambda}$ to have total mass 1, the characters form an orthonormal basis for $L^2(\Gamma/\Lambda)$. Thus, $\{\overline{\phi_h} : h \in H\}$ is an orthonormal basis for $L^2(\Omega)$.

For $f \in PW_{\Omega_0}$, $\widehat{f} \in L^2(\Omega_0) \subset L^2(\Omega)$. Thus,

$$\widehat{f}(\omega) = \sum_{h \in H} \left(\int_\Omega \widehat{f}(\gamma)\phi_h(\gamma)d\widetilde{\mu}_\Gamma(\gamma) \right) \overline{\phi_h}(\omega)$$

$$= c_H \sum_{h \in H} \left(\int_\Omega \widehat{f}(\gamma)e_h(\gamma)d\gamma \right) \overline{\phi_h}(\omega)$$

$$= c_H \sum_{h \in H} f(h)\overline{\phi_h}(\omega) \tag{2.7}$$

in $L^2(\Omega)$.

Multiplying both sides of (2.7) by $\widehat{k}$, we have

$$\widehat{f}(\omega) = \widehat{k}(\omega)\widehat{f}(\omega) = c_H \sum_{h \in H} f(h)\widehat{k}(\omega)\overline{e_h}(\omega) \tag{2.8}$$

in $L^2(\Gamma)$. By the Plancherel Theorem, we may take the inverse Fourier transform of both sides, giving

$$f(x) = c_H \sum_{h \in H} f(h)k(x - h) \tag{2.9}$$

in $L^2(G)$.

To prove the series converges to f uniformly on G, note it follows from the Inversion Theorem that every $f \in PW_{\Omega_0}$ is equal almost everywhere to a continuous function. Hence, without loss of generality, we may assume f is continuous. Similarly, without loss of generality, we may assume k is continuous as well.

For each finite $H_N \subset H$, we have

$$\left| f(x) - c_H \sum_{h \in H_N} f(h)k(x-h) \right| = \left| \int_\Omega \left(\widehat{f}(\gamma) - c_H \sum_{h \in H_N} f(h)\overline{e_h}(\gamma)\widehat{k}(\gamma) \right) e_x(\gamma)d\gamma \right|$$

$$\leq \frac{1}{\sqrt{c_H}} \left\| \widehat{f}(\gamma) - c_H \sum_{h \in H_N} f(h)\overline{e_h}(\gamma)\widehat{k}(\gamma) \right\|_{L^2(\Omega)}.$$

$$(2.10)$$

Therefore, uniform convergence on G follows from the convergence of (2.10) in $L^2(\Omega)$ as $H_N \to H$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

2.1.2 Smooth Sampling

The slow rate of decay of k_Ω motivates our desire for more choices in the selection of a sampling kernel. For example, let $AQ_0 = \Omega \subset \mathbb{R}^d$ be a fundamental cell for Λ. Then

$$k_\Omega(x) = \int_\Omega e_x(\gamma)d\gamma$$

$$= \det(A) \int_{C_0} e_{A^*x}(\gamma)\,d\gamma$$

$$= \det(A) \prod_{j=1}^d \frac{\sin(\pi\langle A_j, x\rangle)}{\pi\langle A_j, x\rangle}, \qquad\qquad (2.11)$$

18

where A_j is the j-th column of A. The best we can say is that k_Ω decays as

$$|k_\Omega(x)| \le \frac{c}{\|x\|_2}.$$

Slow decay rates lead to large truncation errors, a problem in applications where finite approximations to (2.5) are necessary.

This problem is remedied by oversampling. It allows us to choose k such that $\widehat{k}$ is smooth. The smoothness of $\widehat{k}$ significantly increases the rate of decay for k, yielding

$$|k(x)| \le \frac{c}{(1 + \|x\|_2)^p} \quad \forall p \ge 1.$$

Theorem 2.1.5. *Let $\Omega_0 \subset \mathbb{R}^d$ be a bounded, measurable set. Let H and Λ be a pair of dual lattices such that the canonical mapping $\pi : \mathbb{R}^d \to \mathbb{R}^d/\Lambda$ is injective on $\overline{\Omega}_0$. Then there is an $k \in S(\mathbb{R}^d)$, the space of Schwartz functions, such that*

$$\forall f \in PW_{\Omega_0}, \ f(x) = c_H \sum f(h)k(x - h), \tag{2.12}$$

with convergence in $L^2(\mathbb{R}^d)$ and uniformly on $\mathbb{R}^d$.

Proof. By Theorem 2.1.4, it suffices to show that there is an $k \in S(\mathbb{R}^d)$ such that $\widehat{k}\big|_{\Omega_0} \equiv 1$ and π is injective on $\mathrm{supp}(\widehat{k})$.

Since π is injective on $\overline{\Omega}_0$, there is a coset representation Ω such that $\overline{\Omega}_0 \subset \Omega^\circ$. Then

$$d(\Omega_0, \Omega) = \varepsilon > 0.$$

Let $\delta < \varepsilon/2$ and choose $\phi \in C_c^\infty(\mathbb{R}^d)$ such that $\mathrm{supp}(\phi) = B_\delta(0)$ and

$$\int \phi(x)dx = 1.$$

19

We let

$$B_\delta(\Omega_0) = \bigcup_{\omega \in \Omega_0} B_\delta(\omega)$$

and define K to be the convolution of ϕ and $\mathbb{1}_{B_\delta(\Omega_0)}$. Since ϕ is smooth, K is smooth.

For $\omega \in \Omega_0$, $B_\delta(\omega) = \omega - B_\delta(0) \subset B_\delta(\Omega_0)$. Thus

$$K(\omega) = \phi * \mathbb{1}_{B_\delta(\Omega_0)}(\omega)$$

$$= \int_{B_\delta(\Omega_0)} \phi(\omega - \gamma) \, d\gamma$$

$$= \int_{B_\delta(0)} \phi(\gamma) \, d\gamma = 1.$$

Since $d(\Omega, \Omega_0) = \varepsilon > 2\delta$, for $\omega \notin \Omega$, $B_\delta(\omega) \cap B_\delta(\Omega_0) = \emptyset$. This implies,

$$K(\omega) = \int_{B_\delta(\Omega_0)} \phi(\omega - \gamma) d\gamma = 0.$$

Thus,

$$K(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega_0 \\ 0 & \text{if } \omega \notin \Omega \end{cases}.$$

Finally, take k to be the inverse Fourier transform of K. The space of Schwartz functions is invariant under the Fourier transform. Since $K \in C_c^\infty(\mathbb{R}^d) \subset S(\mathbb{R}^d)$, $k \in S(\mathbb{R}^d)$. $\qquad \square$

2.2   Non-Uniform Sampling and Fourier Frames

In non-uniform sampling, lattices are replaced with sampling sets.

Definition 2.2.1 (Sampling Sets)**.** Let $\Omega \subset \Gamma$ have finite Haar measure. A set

$S \subset G$ is said to be a *sampling set* for Ω if $\exists A, B > 0$ such that,

$$\forall f \in PW_\Omega, \quad A\|f\|^2 \leq \sum_{s \in S} |f(s)|^2 \leq B\|f\|^2. \tag{2.13}$$

By (2.13), a bandlimited function is uniquely determined by its values taken over a sampling set. In fact, the definition gives a little more: If S is a sampling set for PW_Ω, then Ω-bandlimited functions can be reconstructed from their sampled values.

The Plancherel Theorem, the Inversion Theorem, and the fact that $L^2(\Omega) \subset L^1(\Omega)$ for sets of finite measure, together prove that S is a sampling set for PW_Ω if and only if its reflection, $S^- = \{-s : s \in S\}$, generates a Fourier frame for $L^2(\Omega)$. Since frame spectra are invariant under reflection, S is a sampling set if and only if it is a spectrum.

Frame expansion provides a reconstruction formula for $f \in PW_\Omega$ from its sampled values. If S is a sampling set and $\mathcal{F}$ is the frame operator associated with the Fourier frame generated by S^-, taking the inverse Fourier transform of the dual frame expansion gives

$$f(x) = \sum_{s \in S} f(s) k_s(x), \ \widehat{k}_s(x) = \mathcal{F}^{-1}(\overline{e_s}(x)). \tag{2.14}$$

The reconstruction converges in $L^2(G)$.

2.2.1   Beurling Density in $\mathbb{R}^d$

Landau proved that the Nyquist condition, necessary for uniform sampling, generalizes to a necessary condition for non-uniform sampling in $\mathbb{R}^d$ [37].

Definition 2.2.2 (Beurling Density). For $r > 0$, let $Q_r(x)$ denote the cube of sidelength r centered at x. The *upper and lower Beurling densities* of S are given by

$$D^+(S) = \limsup_{r \to \infty} \sup_{x \in \mathbb{R}^d} \frac{\text{card}(S \cap Q_r(x))}{r^d} \tag{2.15}$$

$$D^-(S) = \liminf_{r \to \infty} \inf_{x \in \mathbb{R}^d} \frac{\text{card}(S \cap Q_r(x))}{r^d}. \tag{2.16}$$

A set $S \subset \mathbb{R}^d$ is *uniformly discrete* if there is a neighborhood of the origin, U, such that

$$\forall s \in S, \ (s + U) \cap S = \{s\};$$

S is said to be *relatively separated* if it is the union of finitely many uniformly discrete sequences.

Theorem 2.2.1 (Landau, [37]). *Let $\Omega \subset \mathbb{R}^d$ be a set of positive measure. If a uniformly discrete S is a sampling set for PW_Ω, then $D^-(S) \geq m(\Omega)$.*

This statement of Landau's theorem differs from his original statement by a factor of $(2\pi)^d$ because of our normalization of the Fourier transform.

In general, the Nyquist condition is not strong enough to guarantee sampling, even when S is a lattice. When $\Omega \subset \mathbb{R}$ is an interval, however, Beurling proved that the density condition with strict inequality is sufficient to guarantee that a uniformly discrete S is a sampling set for Ω [6].

Seip and Jaffard showed, independently, that a sampling set must necessarily be relatively separated. Using this fact, the condition that S be uniformly discrete may be removed from Beurling's and Landau's results [29, 51].

Seip and Ortega-Cerdà solved the remaining, difficult problem of characterizing the sampling sets with $D^-(S) = m(\Omega)$ [45]. They characterized such sampling sets as the zero sets of a certain family of entire functions.

Landau proved Theorem 2.2.1 via an investigation of the eigenvalues of the bounded, self-adjoint positive operator $P_\Omega Q_U P_\Omega$ defined on $L^2(\mathbb{R}^d)$, where P_Ω is orthogonal projection onto PW_Ω and Q_U is orthogonal projection onto $L^2(U)$. A shorter, yet somewhat less precise, proof was provided in [21] based on an adaptation of the Ramanathan-Steger Comparison Principle [48].

Theorem 2.2.2 (Comparison Principle, [21])**.** *Let*

$$F = \{f_k : 1 \le k \le n\}, \ H = \{h_j : 1 \le j \le m\} \subset PW_\Omega$$

and suppose that $\widehat{H} \subset L^\infty(\Omega)$. *Let*

$$S = \bigcup_{k=1}^{n} S_k \ and \ T = \bigcup_{j=1}^{m} T_j.$$

Assume that $\{f(x - s) : \ f \in F, s \in S\}$ *is a Riesz basis for its closure in* PW_Ω *and that* $\{h(x - t) : \ h \in H, t \in T\}$ *is a frame for* PW_Ω. *Then,* $\forall \varepsilon > 0, \exists R > 0$ *such that* $\forall r \ge 0$ *and* $\forall y \in \mathbb{R}^d$,

$$(1 - \varepsilon) \sum_{k=1}^{n} \mathrm{card}(S_k \cap B_r(y)) \le \sum_{j=1}^{m} \mathrm{card}(T_j \cap B_{r+R}(y)). \qquad (2.17)$$

2.2.2   Beurling Density in LCAGs

Landau's theorem can be generalized to compactly generated, locally compact abelian groups [22]. The first step is to find a suitable notion of density in this

setting. Motivated by (2.17), the authors of [22] define the density of a discrete sequence by comparing it to some fixed lattice.

Let G and Γ be a dual pair of LCAGs such that Γ is compactly generated. By Theorem 2.1.2, $\Gamma = \mathbb{R}^d \times \mathbb{Z}^n \times \Pi$, where Π is a compact abelian group. Hence, $G = \mathbb{R}^d \times \mathbb{T}^n \times D$, where D is a discrete (possibly uncountable) abelian group. The *fundamental lattice* in G is chosen to be $H_0 = \mathbb{Z}^d \times \{0\}^n \times D$.

Definition 2.2.3. Let $a, b > 0$. We write $aS \preceq bT$ if, given $\varepsilon > 0$, $\exists$ a compact $K \subset G$ such that, $\forall$ compact $L \subset G$,

$$(1 - \varepsilon)a \operatorname{card}(S \cap L) \leq b \operatorname{card}(T \cap (K + L)). \tag{2.18}$$

Definition 2.2.4 (Beurling Density)**.** Let S be a discrete subset of G. The upper and lower Beurling densities of S are given by

$$D^+(S) = \inf\{b : \ S \preceq bH_0\} \tag{2.19}$$

$$D^-(S) = \sup\{a : \ aH_0 \preceq S\}. \tag{2.20}$$

To get a feel for 2.19 and 2.20, we provide the proof of a result from [22] showing that this definition is in fact a generalization of Beurling density on $\mathbb{R}^d$.

Proposition 2.2.1. *When $G = \mathbb{R}^d$ the two notions of Beurling density agree for all discrete subsets S.*

Proof. We will show that both definitions of lower Beurling density are equivalent. The equivalence of upper Beurling density can be proved similarly.

Let $S \subset \mathbb{R}^d$ and let

$$a_0 = \sup\{a : \ a\mathbb{Z}^d \preceq S\}.$$

We want to show

$$a_0 \leq \liminf_{r \to \infty} \inf_{x \in \mathbb{R}^d} \frac{\mathrm{card}(S \cap Q_r(x))}{r^d}.$$

Without loss of generality, we may assume $a \neq 0$.

Given $a < a_0$, for every $\varepsilon \in (0, 1)$, there is a compact $K \subset \mathbb{R}^d$ satisfying (2.18). Since K is compact, there is an $R > 0$ such that $K \subseteq Q_R(0)$. Let $L = Q_r(x)$ for some $x \in \mathbb{R}^d$ and $r > 0$. Then $L + K \subset Q_{R+r}(x)$ and by (2.18),

$$\mathrm{card}(Q_{R+r}(x) \cap S) \geq (1 - \varepsilon)a\, \mathrm{card}(Q_r(x) \cap \mathbb{Z}^d)$$

$$\geq (1 - \varepsilon)a(r - 1)^d$$

$$= (1 - \varepsilon)a \left(\frac{r - 1}{r + R} \right)^d (r + R)^d.$$

Thus, for sufficiently large r,

$$\mathrm{card}(Q_{R+r}(x) \cap S) \geq (1 - \varepsilon)^2 a (r + R)^d.$$

This is true for all $\varepsilon \in (0, 1)$ and $a < a_0$, hence

$$\liminf_{r \to \infty} \inf_{x \in \mathbb{R}^d} \frac{\mathrm{card}(S \cap Q_r(x))}{r^d} \geq a_0.$$

On the other hand, suppose

$$\liminf_{r \to \infty} \inf_{x \in \mathbb{R}^d} \frac{\mathrm{card}(S \cap Q_r(x))}{r^d} = a_0.$$

Again, without less of generality, we may assume $a_0 \neq 0$. Then, for every $\varepsilon \in (0, 1)$, there exists an $R > 0$ such that

$$\forall r \geq R, \quad \mathrm{card}(Q_r(x) \cap S) \geq (1 - \varepsilon)a_0 r^d.$$

Let C_0 be a tile for the lattice $R\mathbb{Z}^d$ such that $\overline{C_0} = Q_R(0)$. Let L be a compact set; L intersects finitely many translates of C_0 by elements of $R\mathbb{Z}^d$. Write these as $C_1, \ldots, C_N$. Then

$$L = \bigcup_{n=1}^{N} (C_n \cap L) \subseteq \bigcup_{n=1}^{N} C_n,$$

with each union being pairwise disjoint.

Each $x \in C_n$ satisfies $\|l - x\|_\infty \leq R$ for some $l \in L$. Thus

$$\bigcup_{n=1}^{N} C_n \subseteq L + Q_{2R}(0).$$

Then, we have

$$\mathrm{card}(S \cap (L + Q_{2R}(0))) \geq \mathrm{card}(S \cap (\bigcup_{n=1}^{N} C_n))$$

$$= \sum_{n=1}^{N} \mathrm{card}(S \cap C_n)$$

$$\geq (1 - \varepsilon) a_0 N R^d.$$

Moreover, if we assume $R \in \mathbb{N}$,

$$\mathrm{card}(\mathbb{Z}^d \cap L) \leq \mathrm{card}(\mathbb{Z}^d \cap (\bigcup_{n=1}^{N} C_n))$$

$$= \sum_{n=1}^{N} \mathrm{card}(\mathbb{Z}^d \cap C_n)$$

$$= N R^d.$$

Hence, $a_0 \mathbb{Z}^d \preceq H$. $\qquad\square$

2.2.3 Density Conditions on LCAGs

An argument based on a further generalization of the comparison principle can then be used to prove the Nyquist condition in the setting of compactly generated,

locally compact abelian groups.

Theorem 2.2.3 (Gröchenig, Kutyniok, and Seip, [22])**.** *Let* $\Omega \subset \Gamma$ *be a set of positive measure and assume that* μ_Γ *has been normalized so that* $\mu_\Gamma([-\pi, \pi]^d \times \{0\} \times \Pi) = 1$. *If* S *is a sampling set for* PW_Ω, *then* $D^-(\Omega) \geq \mu_\Gamma(\Omega)$.

They also provide a characterization of sets S such that the upper inequality of (2.13) is satisfied, i.e. sets S that are a Bessel spectrum for Ω. The result is a nice generalization of what is known for $\mathbb{R}^d$.

Theorem 2.2.4 ([Gröchenig, Kutyniok, and Seip, [22])**.** *A set* S *is a Bessel spectrum for a set of positive measure if and only if it is relatively separated.*

We recast this as a density condition below.

Theorem 2.2.5. *Let* $\Omega \subset \Gamma$ *be a set of positive, finite measure. Then* S *is a Bessel spectrum for* Ω *if and only if* $D^+(S) < \infty$.

To prove Theorem 2.2.5, it suffices to prove the following.

Proposition 2.2.2. $D^+(S) < \infty$ *if and only if* S *is relatively separated.*

Proof. In the following, let H_0 be the fundamental lattice in G. Let C_0 be a tile for H_0 which contains the identity. The translates of C_0 by the elements, h, of the fundamental lattice will be written as C_h.

($\implies$) Since $D^+(S) < \infty$, there is an $N > 0$ and a compact set K such that, for every $h \in H$,

$$\mathrm{card}(S \cap C_h) \leq N \, \mathrm{card}(H_0 \cap (K + C_h)).$$

However,

$$\mathrm{card}(H_0 \cap (K + C_h)) = \mathrm{card}(H_0 \cap (K + C_0)).$$

Hence, there is an $M > 0$ such that each C_h contains at most M points. This implies that we can write S as the disjoint union of M uniformly discrete sequences.

($\Longleftarrow$) Since S is relatively separated, there is an $N > 0$ such that each C_h contains at most N points of S.

Let $K = \overline{C_0 \cup C_0^-}$ and let L be any compact subset of G. There exist $N_L < \infty$ tiles such that

$$L \subset \bigcup_{n=1}^{N_L} C_{h_n}.$$

It follows that

$$\mathrm{card}(S \cap L) \le N_L \cdot N.$$

Furthermore,

$$L \cap C_{h_n} \ne \emptyset \implies h_n \in K + L.$$

Hence,

$$\mathrm{card}(H_0 \cap (K + L)) \ge N_L.$$

Therefore, $D^+(S) \le N < \infty$. $\qquad \square$

2.2.4 Stability of Sampling Sets

For a relatively compact $\Omega \subset \mathbb{R}^d$, the stability of Bessel spectra under bounded perturbations and the stability of sampling sets (or frame spectra) under small perturbations is well known (see, e.g., [12]). We adapt an argument from [22] to establish these facts for compactly generated, locally compact abelian groups.

Given a function $g : G \rightarrow \mathbb{C}$ and a set $U \subset G$, define

$$g^\star(x) = \sup_{u \in U} |g(x + u)|.$$

As in [22], we will use the following fact from [49]. Given a relatively compact $\Omega \subset \Gamma$ there is a $g \in L^1(G)$ such that:

- $\widehat{g}(\omega) = 1, \ \forall \omega \in \Omega$, and

- for any compact, symmetric neighborhood of the origin, U,

$$g^\star \in L^1(G). \tag{2.21}$$

Theorem 2.2.6. *Let $\Omega \subset \Gamma$ be a set of positive, finite measure and let $S = \{s_j\}$ be a Bessel spectrum for Ω. Let $V \subset G$ be a relatively compact, symmetric neighborhood of the identity. If $T = \{t_j\}$ satisfies,*

$$t_j \in s_j + V$$

for each j in the indexing set, J, then T is also a Bessel spectrum for Ω.

Proof. If $\exists C > 0$ such that $\forall f \in PW_\Omega$,

$$\sum_{j \in J} |f(s_j) - f(t_j)|^2 \leq C\|f\|^2 \tag{2.22}$$

then

$$\sum_{j \in J} |f(t_j)|^2 = \sum_{j \in J} |f(t_j) - f(s_j) + f(s_j)|^2 \leq (\sqrt{B} + \sqrt{C})^2 \|f\|^2$$

where B is the Bessel bound for S. Hence it suffices to prove (2.22).

29

Let $g \in L^1(G)$ be as in (2.21), and let $g^\star$ be defined using $U = \overline{V + V}$. Then, for each j, we have

$$|f(t_j) - f(s_j)|^2 = |f * g(t_j) - f * g(s_j)|^2$$

$$= \left| \int_G f(x) \left(g(t_j - x) - g(s_j - x) \right) dx \right|^2$$

$$\leq \int_G |f(x)|^2 |g(t_j - x) - g(s_j - x)| dx \int_G |g(t_j - y) - g(s_j - y)| dy$$

$$\leq 2\|g\|_1 \int_G |f(x)|^2 |g(t_j - x) - g(s_j - x)| dx$$

We will show that

$$\sum_{j \in J} |g(t_j - x) - g(s_j - x)| \in L^\infty(G).$$

For a fixed $x_0 \in G$, let $V_j = V + s_j - x_0$. Then

$$\sum_{j \in J} |g(t_j - x_0) - g(s_j - x_0)| = \sum_{j \in J} (m_G(V))^{-1} \int_{V_j} |g(t_j - x_0) - g(s_j - x_0)| dx$$

$$\leq 2(m_G(V))^{-1} \sum_{j \in J} \int_{V_j} |g^\star(x)| dx$$

$$= 2(m_G(V))^{-1} \sum_{j \in J} \sum_{h \in H_0} \int_{C_h \cap V_j} |g^\star(x)| dx$$

$$= 2(m_G(V))^{-1} \sum_{h \in H_0} \sum_{j \in J} \int_{C_h \cap V_j} |g^\star(x)| dx \qquad (2.23)$$

where H_0 is the fundamental lattice, C_0 is a fixed tile for H_0, and $C_h = h + C_0$.

We claim that there is an $N \in \mathbb{N}$ such that at most N of the V_j intersect C_h for each h. Note that

$$C_h \cap V_j \neq \emptyset \implies s_j \in C_h + V_0,$$

where $V_0 = V + x_0$. We have assumed that S is a Bessel spectrum and so $D^+(S) < \infty$.

30

Hence, there is an $N > 0$ and a compact set K such that

$$\mathrm{card}(S \cap (V_0 + C_h)) \leq N \, \mathrm{card}(H_0 \cap (V_0 + C_h + K))$$

$$= N \, \mathrm{card}(H_0 \cap (V_0 + C_0 + K)).$$

The set $V_0 + C_0 + K$ is compact and thus contains only finitely many $h \in H_0$. Hence,

$$\mathrm{card}(S \cap (V_0 + C_h)) \leq M < \infty.$$

where M is independent of h. This, along with (2.23), implies that

$$\sum_{j \in J} |g(t_j - x_0) - g(s_j - x_0)| \leq 2(m_G(V))^{-1} \sum_{h \in H_0} M \int_{C_h} |g^\star(x)| dx$$

$$\leq 2M(m_G(V))^{-1} \|g^\star\|_1$$

Therefore,

$$\sum_{j \in J} |f(s_j) - f(t_j)|^2 \leq 4M(m_G(V))^{-1} \|g^\star\|_1 \|g\|_1 \|f\|^2.$$

$\square$

Theorem 2.2.7. *Let $\Omega \subset \Gamma$ be a set of positive, finite measure and let $S = \{s_j\}$ be a sampling set for PW_Ω. There exists a relatively compact, symmetric neighborhood of the identity, $V \subset G$, such that, if $T = \{t_j\}$ satisfies*

$$t_j \in s_j + V,$$

for every j in the indexing set, J, then T is also a sampling set for PW_Ω.

Proof. By Theorem 2.2.6, any such T will satisfy the Bessel bound. If $\varepsilon < A$ and

$$\sum_{j \in J} |f(s_j) - f(t_j)|^2 \leq \varepsilon \|f\|^2 \tag{2.24}$$

then

$$\left(\sum_{j \in J} |f(t_j)|^2\right)^{1/2} \geq \left(\sum_{j \in J} |f(s_j)|^2\right)^{1/2} - \left(\sum_{j \in J} |f(s_j) - f(t_j)|^2\right)^{1/2}$$

$$\geq (\sqrt{A} - \sqrt{\varepsilon})\|f\|.$$

Squaring both sides gives the lower sampling inequality. Hence, it suffices to prove (2.24).

Let g be defined as in (2.21). Then, $\forall f \in PW_\Omega$,

$$|f(t_j) - f(s_j)|^2 = |f * g(t_j) - f * g(s_j)|^2$$

$$= |\int_G f(x) \left(g(t_j - x) - g(s_j - x)\right) dx|^2$$

$$\leq \int_G |f(x)|^2 |g(t_j - x) - g(s_j - x)| dx \int_G |g(t_j - y) - g(s_j - y)| dy$$

$$\leq \int_G |f(x)|^2 |g(t_j - x) - g(s_j - x)| dx \int_G |g(y) - g(v_j + y)| dy$$

where $v_j = s_j - t_j \in V$. Given $\delta > 0$, choose V small enough so that

$$\int_G |g(y) - g(v_j + y)| dy < \delta.$$

Since S is a sampling set, it is relatively separated. Hence,

$$S = \bigcup_{n=1}^N S_n,$$

where the union is pairwise disjoint and each $S_n = \{s_i^n\}$ is uniformly discrete. Making V smaller if necessary, choose $W \supset V$ small enough so that

- $\{W_i = W + s_i^1\}$ is pairwise disjoint,

- $\text{card}(S \cap W_i) \leq N$, and

- for each $s_j \in W_i$,

$$V + s_j \subset W_i.$$

Fix $x_0 \in G$. Let $W_i^0 = W + s_i^1 + x_0$ and let $g^\star$ be defined using $U = \overline{W + W}$.

Then,

$$\sum_{i \in J} \sum_{s_j \in W_i} |g(t_j - x_0) - g(s_j - x_0)|$$

$$= (m_G(W))^{-1} \sum_{i \in J} \sum_{s_j \in W_i} \int_{W_i^0} |g(t_j - x_0) - g(s_j - x_0)| dx$$

$$\leq 2N(m_G(W))^{-1} \sum_{i \in J} \int_{W_i^0} |g^\star(x)| dx$$

$$\leq 2N(m_G(W))^{-1} \|g^\star\|_1.$$

Thus,

$$\sum_{j \in J} |f(s_j) - f(t_j)|^2 \leq 2N\delta(m_G(W))^{-1} \|g^\star\|_1 \|f\|^2.$$

Since W, and hence $g^\star$, depends only on S, V can be chosen so that

$$2N\delta(m_G(W))^{-1} \|g^\star\|_1 \leq \varepsilon.$$

$\square$

Chapter 3

Fourier Spectra for $L^2(\mu)$

3.1   General Properties

A natural generalization of the local Fourier analysis discussed in Chapter 2 is

the Fourier analysis for $L^2(\mu)$, where μ is an arbitrary measure in $M_1(\Gamma)$. Unlike the

case for sets of finite measure, there is no guarantee in this general setting that such

an analysis can be carried out. A fundamental problem is to classify the measures

which admit Fourier frames. Ultimately, the goal is to develop a set of criteria so

that, given $\mu \in M_1(\Gamma)$, we not only know if μ is spectral, but we can completely

characterize the spectra it admits. However, even in the simplest case, when $\Gamma = \mathbb{R}$,

such problems are far from being resolved.

We will investigate the role the geometry of the support of μ plays in determin-

ing its spectra. The degree to which geometry influences spectral properties clearly

decreases with the rigidity of the requirements of the type in question. Though, as

we will see in Section 3.3.1, it manifests itself even on the level of Bessel spectra.

For Fourier spectra, this study is closely related to the Fuglede conjecture.

Fuglede's Conjecture ([20]). *A set $\Omega \subset \mathbb{R}^d$ of finite measure admits a Fourier*

spectrum S if and only if Ω tiles $\mathbb{R}^d$, that is, if and only if there is a $T \subset \mathbb{R}^d$ such

that

$$\mathbb{R}^d = \bigcup_{t \in T} \Omega + t$$

and the union is pairwise disjoint almost everywhere.

This question arose naturally out of Fuglede's work on differential operators. He was interested in classifying the domains $\Omega \subset \mathbb{R}^d$ such that the family of differential operators, $\Delta_k = -i \frac{\partial}{\partial x_k}$, can be extended to a family of commuting self-adjoint operators acting on $L^2(\Omega)$. He proved that a domain admits such an extension if and only if $L^2(\Omega)$ contains a Fourier basis. The conjecture was posited based on his observation of the relationship between Fourier sets and tilings when either T or S is assumed to be a lattice.

A number of positive results were achieved—

- the conjecture is true for convex planar domains [27];

- convex sets having a smooth boundary are not Fourier sets (a positive result since it also known that such sets do not tile) [28];

- the conjecture is true for the union of two intervals [35];

—until Tao found a counterexample in $d \geq 5$ with a finite configuration of cubes that has a Fourier basis, but does not tile [52]. A similar construction was used in [43] and [33] to prove that the same implication is false in $d = 4$ and $d = 3$, respectively. A counterexample to the converse was similarly found in a string of papers for $d = 5$, $d = 4$, and $d = 3$ [33, 18, 17]. Both directions remain open for $d = 1$ and $d = 2$.

The problem of characterizing sets which admit Riesz spectra has a similar type of geometric rigidity, however, it has proved to be more difficult than its Fourier counterpart. In $\mathbb{R}$, there are several families of finite unions of intervals that are known to have Riesz spectra (see [41, 39] for example). It is still an open question whether any finite union of intervals admits a Riesz spectra. In $\mathbb{R}^2$, it is known that convex polygons are Riesz spectral sets [40]. Beyond these results, not much else is known. The machinery involved with Riesz spectra is significantly more technically difficult than that for Fourier spectra and presents an obstacle that must be overcome.

For Fourier frames in general, the only case that has been completely solved is a single interval in $\mathbb{R}$. In all dimensions and in LCAGs, we have seen that the size of the set determines a necessary condition on the density of its spectra, however this condition is far from being sufficient. On the other hand, recent results have also shown that frames are significantly less rigid than their basis counterparts. In fact, there exist families of *universal spectra*:

Theorem 3.1.1 (Matei and Meyer, [42]). *There exists sets S such that $D^-(S) > m(\Omega)$ is sufficient for S to be a frame spectrum for any Ω such that $m(\partial\Omega) = 0$.*

In this section, our investigation of the geometry of $\text{supp}(\mu)$ will focus on *size*. The effect of the size of the support of μ on its spectral properties sheds light on the fundamental problem posed above. Namely, we use necessary density conditions imposed by the size of $\text{supp}(\mu)$ in the setting of LCAGs to generalize several recent results addressing the problem of characterizing measures of a given spectral type.

36

3.1.1 Necessary Density Conditions

When Γ is not discrete, each measure $\mu \in M_1(\Gamma)$ can be written as $\mu = \mu_d + \mu_{sc} + \mu_{ac}$ where μ_d is discrete and μ_{sc} and μ_{ac} are singularly continuous and absolutely continuous with respect to Haar measure. In [24], it was shown that when $\Gamma = \mathbb{R}^d$, if $L^2(\mu)$ contains a Fourier frame then μ decomposes purely as one of these three types. We extend this result to compactly generated LCAGs.

First, we establish necessary density conditions for measures of each type.

Lemma 3.1.1. *If S is a Bessel spectrum for $\mu \in M_1(\Gamma)$, then*

$$D^+(S) < \infty.$$

Proof. For $\mu \in M_1(\Gamma)$, $\widehat{\mu}(0) = 1$. Since $\widehat{\mu}$ is continuous, given $c \in (0,1)$, there is an open neighborhood U of 0 such that

$$\forall x \in U, \quad \widehat{\mu}(x) \geq c > 0.$$

Let $B > 0$ be the Bessel bound for $E(S)$. For every $x \in G$,

$$B \geq \sum_{s \in S} |\widehat{\mu}(s-x)|^2 \geq \sum_{s \in (U+x)} |\widehat{\mu}(s-x)|^2 \geq c \cdot \mathrm{card}(S \cap (U+x)).$$

Hence, there is an $N \in \mathbb{N}$ such that every translate of U contains at most N elements of S.

Let H_0 be the fundamental lattice in G. Let C_0 be a tile for H_0 which contains the identity. Since $\overline{C_0}$ is compact, there is an $N_0 \in \mathbb{N}$ such that C_0 can be covered by N_0 translates of U. Moreover, each tile $C_h = C_0 + h$ can also be covered by N_0 translates of U. Hence, each C_h contains at most NN_0 points of S. This implies S is relatively separated. By Proposition 2.2.2, $D^+(S) < \infty$. $\square$

Lemma 3.1.2. *If S is a spectrum for a discrete measure μ, then*

$$\operatorname{card}(S) < \infty.$$

Proof. We have $\mu = \sum_{\lambda \in \Lambda} c_\lambda \delta_\lambda$, where Λ is a discrete subset of Γ. Let $\phi(\gamma) = \mathbb{1}_{\{\lambda_0\}}(\gamma)$ for some $\lambda_0 \in \Lambda$. Then

$$\operatorname{card}(S)|c_{\lambda_0}|^2 = \sum_{s \in S} |c_{\lambda_0} e_s(\lambda_0)|^2$$

$$= \sum_{s \in S} |\langle \phi, e_s \rangle|^2 < \infty.$$

$\square$

Definition 3.1.1. We say that $f : G \to \mathbb{C}$ is *locally integrable* on G if $f(x)\mathbb{1}_K(x) \in L^1(G)$ for every compact $K \subset G$.

Proposition 3.1.3. *Let $f : G \to \mathbb{C}$ be a Borel measurable function. Then $f \in L^2(G)$ if and only if there is a discrete $S \subset G$ with $D^-(S) > 0$ such that*

$$f_S(x) = \sum_{s \in S} |f(s - x)|^2$$

is locally integrable on G.

Proof. ($\Longrightarrow$) Suppose $f \in L^2(G)$. Let H_0 be the fundamental lattice for G and let C_0 be a tiling cell associated with H_0. Then

$$\|f\|^2 = \sum_{h \in H_0} \int_{C_0 + h} |f(x)|^2 dx$$

$$= \sum_{h \in H_0} \int_{C_0^-} |f(h - y)|^2 dy$$

$$= \int_{h - C_0} f_{H_0}(y) dy, \tag{3.1}$$

38

for any $h \in H_0$.

Since
$$G = \bigcup_{h \in H_0} (h - C_0),$$
given a compact set $K \subset G$, there exist finitely many $h_1, \ldots, h_N \in H_0$ such that
$$K \subset \bigcup_{n=1}^{N} (h_n - C_0).$$

From (3.1), it follows that
$$\int_K f_{H_0}(x) \leq N \|f\|^2.$$

($\Longleftarrow$) Suppose $f_S(x) = \sum_{s \in S} |f(s - x)|^2$ is locally integrable for some discrete set $S \subset G$, with $D^-(S) = D > 0$. Then, given $\epsilon > 0$, there is a compact set K such that for every compact $L \subset G$,

$$\mathrm{card}((K + L) \cap S) \geq D(1 - \epsilon)\, \mathrm{card}(L \cap H_0).$$

In particular, there is a compact set K containing the identity such that for each $h \in H_0$,

$$\mathrm{card}((K + h) \cap S) \geq 1.$$

Let $U = (K \cup K^-) + C_0$, where C_0 is again a tiling cell for H_0. For each $h \in H_0$, there is an $s \in S$ such that $s \in (K \cup K^-) + h$. This implies $h \in (K \cup K^-) + s$. Hence, $h + C_0 \subset s + U$. Since C_0 is a tiling cell,

$$G = \bigcup_{h \in H_0} (C_0 + h) \subset \bigcup_{s \in S} (U + s).$$

However, f_S is locally integrable and $\overline{U}$ is compact. Therefore,

$$\|f\|^2 \leq \sum_{s \in S} \int_{s+U} |f(x)|^2 dx$$

$$\leq \int_{U^-} f_S(x) dx < \infty.$$

$\square$

Lemma 3.1.4. *If S is a spectrum for a singular measure $\mu \in M_1(\Gamma)$, then*

$$D^-(S) = 0.$$

Proof. If S is a spectrum for μ, we have that

$$\widehat{\mu}_S(x) = \sum_{s \in S} |\widehat{\mu}(s - x)|^2 \leq B < \infty$$

for every $x \in G$. Hence, $\widehat{\mu}_S \in L^\infty(G)$ and is locally integrable. By Proposition 3.1.3, if $D^-(S) > 0$, then $\widehat{\mu} \in L^2(G)$. This would imply that $\mu = \phi d\gamma$ where $\phi \in L^2(\Gamma)$; μ is singular, however, so ϕ must be identically 0. This implies $\mu = 0$, a contradiction. $\square$

For each absolutely continuous measure $\mu \in M_1(\Gamma)$, we write ϕ_μ for its Radon-Nikodym derivative. Since we assume μ to be finite, each ϕ_μ is a non-negative function in $L^1(\Gamma)$.

Definition 3.1.2. We define the *nonzero set* of a function $\phi : \Gamma \to \mathbb{C}$ to be

$$\Sigma(\phi) = \{\gamma \in \Gamma : \phi(\gamma) \neq 0\}.$$

Also, given $N \geq 1$, define

$$\Sigma_N(\phi) = \{\gamma \in \Gamma : \frac{1}{N} \leq |\phi(\gamma)| \leq N\}.$$

Lemma 3.1.5. *Let ϕ be a Borel measurable function and let Ω be any set of positive, finite measure contained in $\Sigma_N(\phi)$ for some $N \geq 1$. Let μ be the absolutely continuous measure defined by $d\mu = \phi(\gamma)\mathbb{1}_\Omega(\gamma)d\gamma$.*

For any $S \subset G$, $E(S)$ is a Fourier frame for $L^2(\mu)$ if and only if it is a Fourier frame for $L^2(\Omega)$.

Proof. ($\implies$) Let $A, B > 0$ be the frame bounds for $E(S)$ in $L^2(\mu)$. For each $\psi \in L^2(\Omega)$, define

$$(\psi/\phi)(\gamma) = \begin{cases} \frac{\psi(\gamma)}{\phi(\gamma)} & \text{if } \phi(\gamma) \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Then $\psi/\phi \in L^2(\mu)$, since

$$\|\psi/\phi\|_\mu^2 = \int_\Omega |(\psi/\phi)(\gamma)|^2\phi(\gamma)d\gamma \leq N\|\psi\|^2.$$

Thus,

$$\sum_{s \in S} |\langle \psi, e_s \rangle|^2 = \sum_{s \in S} |\langle \psi/\phi, e_s \rangle_\mu|^2 \leq B\|\psi/\phi\|_\mu^2 \leq BN\|\psi\|^2.$$

Similarly,

$$\sum_{s \in S} |\langle \psi, e_s \rangle|^2 = \sum_{s \in S} |\langle \psi/\phi, e_s \rangle_\mu|^2 \geq A\|\psi/\phi\|_\mu^2 \geq \frac{A}{N}\|\psi\|^2.$$

Therefore, $E(S)$ is a Fourier frame for $L^2(\Omega)$ with frame bounds $A', B' > 0$ satisfying

$$\frac{A}{N} \leq A' \leq B' \leq BN.$$

($\impliedby$) Now let $A, B > 0$ be the frame bounds for $E(S)$ in $L^2(\Omega)$. For each $\psi \in L^2(\mu)$, $\psi\phi \in L^2(\Omega)$, since

$$\|\psi\phi\|^2 = \int_\Omega |\psi(\gamma)\phi(\gamma)|^2d\gamma \leq N\|\psi\|_\mu^2.$$

Thus,

$$\sum_{s \in S} |\langle \psi, e_s \rangle_\mu|^2 = \sum_{s \in S} |\langle \psi\phi, e_s \rangle|^2 \leq B\|\psi\phi\|^2 \leq BN\|\psi\|_\mu^2.$$

Similarly,

$$\sum_{s \in S} |\langle \psi, e_s \rangle_\mu|^2 = \sum_{s \in S} |\langle \psi\phi, e_s \rangle|^2 \geq A\|\psi\phi\|^2 \geq \frac{A}{N}\|\psi\|_\mu^2.$$

Therefore, $E(S)$ is a frame for $L^2(\mu)$ with frame bounds $A', B' > 0$ satisfying

$$\frac{A}{N} \leq A' \leq B' \leq BN.$$

$\square$

Lemma 3.1.6. *If S is a spectrum for an absolutely continuous measure $\mu \in M_1(\Gamma)$,*

then

$$D^-(S) > 0.$$

Proof. Choose $N > 0$ such that $\Sigma_N(\phi_\mu)$ has positive measure. Then for some

compact set $\Omega \subset \Gamma$,

$$0 < \mu_\Gamma(\Omega_N) < \infty, \quad \text{where } \Omega_N = \Omega \cap \Sigma_N(\phi_\mu).$$

By Lemma 3.1.5, H is a frame spectrum for $L^2(\Omega_N)$.

Therefore, by Theorem 2.2.3, we must have that $D^-(S) \geq \mu_\Gamma(\Omega_N) > 0.$ $\square$

Theorem 3.1.2 (Law of Pure Type)**.** *If $L^2(\mu)$ contains a Fourier frame, then μ is*

either discrete, absolutely continuous, or singularly continuous.

Proof. Note that this is trivially true if Γ is discrete. If Γ is not discrete, each

$\mu \in M_1(\Gamma)$ admits a decomposition into its discrete, singularly continuous, and

absolutely continuous parts, written $\mu = \mu_d + \mu_{sc} + \mu_{ac}$. The decomposition is

mutually singular, meaning that Γ can be written as the pairwise disjoint union of measurable sets, Γ_d, Γ_{sc}, and Γ_{ac}, such that each set is non-null only for its corresponding measure.

Thus, the spaces $L^2(\mu_d)$, $L^2(\mu_{sc})$, and $L^2(\mu_{ac})$ can be thought of as closed subspaces of $L^2(\mu)$. This implies that if S is a frame spectrum for $L^2(\mu)$, then it is a frame spectrum for $L^2(\mu_d)$, $L^2(\mu_{sc})$, and $L^2(\mu_{ac})$ simultaneously.

By Lemma 3.1.2, if S is a frame spectrum for $L^2(\mu_d)$ and μ_d is not trivial, then S is finite. This implies that $L^2(\mu)$ is a finite dimensional space, which is true only if μ is discrete. Hence, μ is either purely discrete or purely continuous.

Suppose μ is purely continuous. Lemma 3.1.4 implies that if μ_{sc} is non-trivial, then $D^-(S) = 0$. Lemma 3.1.6 implies that, if μ_{ac} is non-trivial, then $D^-(S) > 0$. Hence, μ is either purely singularly or absolutely continuous. $\qquad\square$

3.1.2 Absolutely Continuous Spectral Measures

Here, we extend the results of [36] to compactly generated LCAGs, providing a complete characterization of compactly supported, absolutely continuous spectral measures in terms of their Radon-Nikodym derivative. We will show that if an absolutely continuous measure is spectral, then its Radon-Nikodym derivative is essentially bounded and cannot decay to zero on the boundary of its nonzero set.

Let $M_a(\Gamma)$ be the space of absolutely continuous measures in $M_1(\Gamma)$.

Lemma 3.1.7. *If $\mu \in M_a(\Gamma)$ is spectral, then*

$$\mu_\Gamma(\Sigma(\phi_\mu)) < \infty.$$

Proof. Note that

$$\Sigma(\phi_\mu) = \bigcup_{N=1}^{\infty} \Sigma_N(\phi_\mu).$$

Suppose $\mu_\Gamma(\Sigma(\phi_\mu)) = \infty$. Given $M > 0$, there is an $N > 0$ and a compact set

$\Omega \subset \Gamma$ such that

$$\mu_\Gamma(\Omega_N) \geq M, \text{ where } \Omega_N = \Omega \cap \Sigma_N.$$

Let $S \subset G$ be a spectrum for μ. By Lemma 3.1.5, $E(S)$ is a frame for $L^2(\Omega_N)$.

Hence, by Theorem 2.2.3,

$$D^-(S) \geq \mu_\Gamma(\Omega_N) \geq M.$$

This holds for arbitrarily large M, hence $D^-(S) = \infty$. This implies $D^+(S) = \infty$,

which contradicts Lemma 3.1.1. $\square$

Lemma 3.1.8. *If $\mu \in M_a(\Gamma)$ is a spectral measure, then there is a constant $c > 0$*

such that

$$\phi_\mu \geq c \text{ a.e. on } \Sigma(\phi_\mu).$$

Proof. For $k \in \mathbb{N}$, let

$$\Omega_k = \{\gamma : \frac{1}{k+1} \leq \phi_\mu(\gamma) < \frac{1}{k}\}.$$

Let $C_h = C_0 + h$ be the translates of a tiling cell C_0 associated with the fundamental

lattice H_0. For each k such that $\mu_\Gamma(\Omega_k) > 0$, choose an $h = h(k)$ such that $\mu_\Gamma(\Omega_k \cap$

$C_h) > 0$. Define

$$\Omega_k^h = \Omega_k \cap C_h.$$

Let S be a spectrum for μ. By Lemma 3.1.1, $D^+(S) < \infty$. Hence, S is a

Bessel spectrum for each C_h. The frame inequality is invariant under spatial and

spectral translations. Thus, the Bessel bound, $B > 0$, for $E(S)$ is the same in each $L^2(C_h)$. Since $\Omega_k^h \subset C_h$, the Bessel bound for each Ω_k^h is independent of k. Letting A be the lower frame bound for $E(S)$ in $L^2(\mu)$, we have

$$
\begin{aligned}
\frac{A}{k+1}\mu_\Gamma(\Omega_k^h) &= \frac{A}{k+1}\int_\Gamma |\mathbb{1}_{\Omega_k^h}(\gamma)|^2 d\mu \\
&\leq \sum_{s\in S}|\int_\Gamma \mathbb{1}_{\Omega_k^h}(\gamma)\overline{e_s}(\gamma)d\mu|^2 \\
&= \sum_{s\in S}|\int_{\Omega_k^h} \phi_\mu(\gamma)\overline{e_s}(\gamma)d\gamma|^2 \\
&\leq B\int_{\Omega_k^h} |\phi_\mu(\gamma)|^2 d\gamma \\
&\leq \frac{B}{k^2}\mu_\Gamma(\Omega_k^h).
\end{aligned}
$$

Rearranging, we have

$$
A \leq \frac{k+1}{k^2}B.
$$

Since B is independent of k, the right hand side tends to 0 as $k \to \infty$. Hence, $\mu_\Gamma(\Omega_k) > 0$ for only finitely many k. $\qquad\square$

Lemma 3.1.9. *Let Ω, Ω' be Borel measurable sets of positive, finite measure. Then $\exists \gamma_0 \in \Gamma$ such that $\mu_\Gamma((\gamma_0 + \Omega) \cap \Omega') > 0$.*

Proof. Consider the convolution

$$
\begin{aligned}
\phi(\gamma_0) &= \mathbb{1}_{\Omega^-} * \mathbb{1}_{\Omega'}(\gamma_0) \\
&= \int \mathbb{1}_{\Omega^-}(\gamma_0 - \gamma)\mathbb{1}_{\Omega'}(\gamma)d\gamma \\
&= \mu_\Gamma((\Omega + \gamma_0) \cap \Omega').
\end{aligned}
$$

45

Note that

$$\widehat{\phi}(0) = \widehat{\mathbb{1}}_{\Omega^-}(0)\widehat{\mathbb{1}}_{\Omega'}(0) = \mu_\Gamma(\Omega)\mu_\Gamma(\Omega') > 0.$$

Since $\widehat{\phi}$ is continuous, it must be greater than zero on a set of positive measure.

Thus, ϕ must be greater than zero on a set of positive measure. In particular, there

is at least one $\gamma_0 \in \Gamma$ such that

$$\phi(\gamma_0) = \mu_\Gamma((\Omega + \gamma_0) \cap \Omega') > 0.$$

$\square$

Lemma 3.1.10. *If $\mu \in M_a(\Gamma)$ is spectral, then $\phi_\mu \in L^\infty(\Gamma)$.*

Proof. For $k \in \mathbb{N}$, let

$$\Omega_k = \{\gamma : k \leq \phi_\mu(\gamma) \leq k+1\}.$$

Choose an $N > 0$ such that $\mu_\Gamma(\Sigma_N(\phi_\mu)) > 0$. If $\mu_\Gamma(\Omega_k) > 0$, by Lemma 3.1.9,

$\exists \gamma_0 \in \Gamma$ such that $\mu_\Gamma(\Omega_k \cap (\Sigma_N + \gamma_0)) > 0$. Define

$$\Omega_k^N = \Omega_k \cap (\Sigma_N + \gamma_0).$$

Let $E(S)$ be a Fourier frame for $L^2(\mu)$ with frame bounds $A, B > 0$. By

Lemma 3.1.5, $E(S)$ is a Fourier frame for $L^2(\Sigma_N)$ with frame bounds $A', B' > 0$

satisfying

$$\frac{A}{N} \leq A' \leq B' \leq BN.$$

Since the frame inequality is invariant under spatial translations and is inherited by

subspaces, $E(S)$ is a Fourier frame with the same frame bounds for $L^2(\Sigma_N + \gamma_0)$,

and hence $L^2(\Omega_k^N)$.

Note $\phi_\mu \mathbb{1}_{\Omega_k^N} \in L^2(\Omega_k^N)$. Thus,

$$\frac{A}{N}k^2\mu_\Gamma(\Omega_k^N) \leq \frac{A}{N}\int\limits_{\Omega_k^N}|\phi_\mu(\gamma)|^2d\gamma$$

$$\leq \sum_{s\in S}|\int\limits_{\Omega_k^N}\phi_\mu(\gamma)\overline{e_s}(\gamma)d\gamma|^2$$

$$= \sum_{s\in S}|\int\limits_{\Gamma}\mathbb{1}_{\Omega_k^N}(\gamma)\overline{e_s}(\gamma)d\mu|^2$$

$$\leq B\int\limits_{\Gamma}|\mathbb{1}_{\Omega_k^N}(\gamma)|^2d\mu$$

$$\leq B(k+1)\mu_\Gamma(\Omega_k^N).$$

Rearranging, we have

$$B \geq \frac{A}{N}\frac{k^2}{k+1}.$$

The right hand side tends to infinity as $k \to \infty$. Therefore, $\mu_\Gamma(\Omega_k) > 0$ for only finitely many k. $\qquad\square$

Theorem 3.1.3. *Let μ be a compactly supported measure in $M_a(\Gamma)$. Then, μ is spectral if and only if there exists an $N > 1$ such that*

$$\mu_\Gamma(\Sigma(\phi_\mu) \setminus \Sigma_N(\phi_\mu)) = 0.$$

Proof. Lemmas 3.1.8 and 3.1.10 together prove that the condition is necessary. To prove that it is sufficient, note that since Γ is compactly generated, we have $\Gamma = \mathbb{R}^d \times \mathbb{Z}^n \times \Pi$, where Π is a compact abelian group. Since $\operatorname{supp}\mu$ is compact, it is contained in $\Omega_N = [-\frac{N}{2}, \frac{N}{2}]^d \times \{-N, \ldots, N\}^n \times \Pi$ for some $N > 0$. Letting $\Lambda_N = (N\mathbb{Z})^d \times (2N\mathbb{Z})^n \times \Pi$, we have that Ω_N is a coset representation of Γ/Λ_N; that is, Ω_N is a tile for the lattice Λ_N.

Therefore, letting $S_N = \Lambda_N^\perp$, $E(S_N)$ is a Fourier basis for $L^2(\Omega_N)$. $\qquad\square$

3.2 Self-Similar Measures

In contrast with absolutely continuous measures, singular measures do not come with many user friendly tools, challenging us to come up with new ideas and a new approach. The class of self-similar singular measures provides a first step in this direction. Derived from iterated function systems, their rich geometric structure gives a tractable approach towards understanding their spectral properties.

In this section, we describe the class of self-similar measures. We show that self-similar measures produce examples of Cantor measures, i.e. measures supported by a Cantor set. We discuss several other constructions of Cantor measures, and show that, for a certain class of self-similar measures, these examples are equivalent.

3.2.1 Invariant Sets and Measures

In [26], Hutchinson studied iterated function systems and their invariants. In particular, he established the idea that each iterated function system induces an invariant set and an invariant measure. His proof of the fact that invariant measures exist and are unique had a gap in it, however. The basic idea was to define a metric on the space of Borel probability measures, $M_1(X)$, and show that each iterated function system induces a contraction with respect to this metric.

Definition 3.2.1 (Hutchinson's metric). A function, ϕ, is *Lipschitz* if its *Lipschitz*

constant

$$r_\phi = \sup_{x,y \in X} \frac{d(\phi(x), \phi(y))}{d(x, y)}$$

is finite. Let $\mathrm{Lip}(X)$ be the space of Lipschitz functions and $\mathrm{Lip}^1(X) \subset \mathrm{Lip}(X)$ be the subspace of functions whose Lipschitz constants are less than or equal to 1. Define

$$\delta(\mu, \nu) = \sup\{|\mu(\phi) - \nu(\phi)| : \phi \in \mathrm{Lip}^1(X) \cap L^1(\mu) \cap L^1(\nu)\}.$$

The existence and uniqueness of invariant measures then follow from the Contraction Mapping Principle if δ is complete. In general, this is not the case.

In [34], the gap is remedied by introducing a new metric on a subspace of $M_1(X)$. We present a simple proof of Hutchinson's original result by showing that, without loss of generality, one can confine the search for an invariant measure within a subspace $M_1(Y) \subset M_1(X)$ where Y is compact subset of X. In this setting, Hutchinson's original proof is correct.

Besides Hutchinson's paper, we found [16] and the last chapter of [19] useful resources in the following.

For convenience, we recall the basic definitions for iterated functions systems from Chapter 1.

Definition 3.2.2 (Iterated Function System). An *iterated function system* (IFS)

$$\Phi = \{\phi_1, \ldots, \phi_N\}$$

is a collection of contraction mappings on a complete metric space X.

Definition 3.2.3 (IFS Operators). An iterated function system induces a mapping

49

on the power set of X, $\mathcal{P}(X)$, given by

$$Y \in P(X) \mapsto \Phi(Y) = \bigcup_{i=1}^{N} Y_i$$

$$Y_i = \phi_i(Y). \tag{3.2}$$

This mapping is called the *Hutchinson operator*. A set is said to be *invariant* with respect to Φ if

$$Y = \Phi(Y) = \bigcup_{i=1}^{N} Y_i.$$

Iterated function systems also induce a mapping on $M(X)$. For $\mu \in M(X)$, the *push forward* of μ by $\phi_i \in \Phi$ is given by

$$\phi_i(\mu)(Y) = \mu(\phi_i^{-1}(Y)). \tag{3.3}$$

Let $c \in \mathbb{R}^N$ be a *probability vector*; that is, for $1 \le i \le N$, $c_i > 0$ and

$$\sum_{i=1}^{N} c_i = 1.$$

For each Φ and c, let $\Phi_c : M_1(X) \to M_1(X)$ be given by

$$\Phi_c(\mu) = \sum_{i=1}^{N} c_i \mu_i.$$

A measure is said to be *invariant* with respect to Φ_c if

$$\mu = \Phi_c(\mu) = \sum_{i=1}^{N} c_i \mu_i.$$

Definition 3.2.4. Since each $\phi_i \in \Phi$ is a contraction, $\phi \in \mathrm{Lip}(X)$ with Lipschitz constant $r_i < 1$. We define

$$r_\Phi = \max_{1 \le i \le N} r_i.$$

50

If Y is invariant, it is invariant under repeated applications of Φ. We will use the following notation for this action:

$$Y = \bigcup_{\iota \in I^n(N)} Y_\iota,$$

where $I^n(N) = \{1, \ldots, N\}^n$, $\iota = \{i_1, \ldots, i_n\}$ and $Y_\iota = \phi_\iota(Y)$, with

$$\phi_\iota = \phi_{i_1} \circ \phi_{i_2} \circ \ldots \circ \phi_{i_n} \in \Phi^n.$$

For completeness, we provide a proof of Hutchinson's theorem on the existence of unique, invariant sets. This particular proof also motivates our patch in the gap of the proof of invariant measures. In the proof, we make use of the following lemma.

Lemma 3.2.1. *Let Y be a closed, bounded set, invariant with respect to Φ. Let $Z \subset X$ be any closed set such that $\Phi(Z) \subset Z$. Then $Y \subset Z$.*

Proof. For any $x \in X$ and $Z \subset X$, let

$$d(x, Z) = \inf_{z \in Z} d(x, Z).$$

In our case, Z is closed, and thus $d(x, Z) = 0$ if and only if $x \in Z$. Define

$$d_Y(Z) = \sup_{y \in Y} d(y, Z).$$

Since Y is bounded, $d_Y(Z) < \infty$.

Given $\varepsilon > 0$, choose an $y_0 \in Y$ such that $d_Y(Z) \leq d(y_0, Z) + \varepsilon$. Since Y is

invariant, there is a $y \in Y$ and a $\phi_i \in \Phi$ such that $y_0 = \phi_i(y)$. Thus we have

$$d_Y(Z) \leq d(\phi_i(y), Z) + \varepsilon$$

$$\leq d(\phi_i(y), Z_i) + \varepsilon$$

$$\leq r_\Phi d(y, Z) + \varepsilon$$

$$\leq r_\Phi d_Y(Z) + \varepsilon.$$

This holds $\forall \varepsilon > 0$, implying

$$d_Y(Z) \leq r_\Phi d_Y(Z).$$

Since $r_\Phi < 1$, we must have $d_Y(Z) = 0$. Therefore, $Y \subset Z$. $\qquad\square$

Note that if Y and Z are both closed, bounded sets, invariant with respect to Φ, we have

$$Y \subset Z \subset Y \implies Z = Y.$$

This proves that if a closed, bounded invariant set exists for Φ, it is unique.

Theorem 3.2.1 (Hutchinson, [26]). *For each iterated function system Φ, there is a unique closed, bounded subset Y such that $\Phi(Y) = Y$. Moreover, Y is compact.*

Proof. We have already proved the uniqueness part of the theorem. We must show that such an invariant set exists.

Since X is complete, each ϕ_i has a unique fixed point $x_i \in X$. Fix a point $x_0 \in X$. Let $R_i = d(x_i, x_0)$ and $R = \max\{R_i : 1 \leq i \leq N\}$. Fix $R_0 \geq \frac{1+r_\Phi}{1-r_\Phi} R$ and let

$X_0 = B_{R_0}(x_0)$. Then, for $1 \le i \le N$ and each $x \in X_0$,

$$d(\phi_i(x), x_0) \le d(\phi_i(x), x_i) + d(x_i, x_0)$$

$$\le r_i d(x, x_i) + R$$

$$\le r_\Phi(R + R_0) + R$$

$$\le R_0.$$

Thus,

$$\Phi(X_0) \subset X_0.$$

This implies that, for $k > 1$,

$$\Phi^k(X_0) \subset \Phi^{k-1}(X_0) \subset \ldots \subset X_0.$$

Define

$$Y = \bigcap_{k=1}^{\infty} \overline{\Phi^k(X_0)}.$$

Let $k \ge 1$ and $\psi \in \Phi^k$. For each $y \in Y$ and $n \ge 1$, there is a sequence $\{x_m^n\} \subset \Phi^n(X_0)$ such that

$$y = \lim_{m \to \infty} x_m^n.$$

Thus,

$$\psi(y) = \psi(\lim_{m \to \infty} x_m^n) = \lim_{m \to \infty} \psi(x_m^n) \in \overline{\Phi^{k+n}(X_0)}.$$

This is true for every $n \ge 1$, hence $\psi(y) \in Y$. By Lemma 3.2.1, Y contains the fixed point for ψ. Thus, $Y \ne \emptyset$ and contains all of the fixed points for the functions in Φ^k for each $k \ge 1$.

Since Y is a closed subset of a complete metric space, it is complete as well. To prove it is compact, we will show it is totally bounded.

Given $\varepsilon > 0$, choose $n \geq 1$ such that $r_\Phi^n R_0 < \varepsilon/4$. Again, we have that each $y \in Y$ is the limit of a sequence $\{x_m^n\} \subset \Phi^n(X_0)$. Choose an m such that $d(y, x_m^n) < \varepsilon/2$; $x_m^n = \psi_m(x_m)$ for some $\psi_m \in \Phi^n$ and $x_m \in X_0$. Let $y_m \in Y$ be the fixed point for ψ_m. Then,

$$d(y, y_m) \leq d(y, x_m^n) + d(x_m^n, y_m)$$

$$\leq \frac{\varepsilon}{2} + r_\Phi^n d(x_m, y_m)$$

$$\leq \frac{\varepsilon}{2} + r_\Phi^n 2R_0$$

$$< \varepsilon.$$

Thus, Y is covered by the collection of ε-balls around each of the fixed points of functions in Φ^n. The covering is finite since there are finitely many functions in Φ^n, each with a unique fixed point. Hence, Y is compact. Note that this argument also implies that Y is the closure of the set of fixed points for the functions $\psi \in \Phi^k$, $k \geq 1$.

To complete the proof, we need to show that $\Phi(Y) = Y$. Clearly, $\Phi(Y) \subset Y$. Let $y \in Y$. There is a sequence of fixed points y_m converging to y. Let $\{\psi_m \in \Phi^{k_m}\}$ (without loss of generality $k_m > 1$) be a sequence of functions corresponding to the fixed points y_m. Each $\psi_m = \phi_{j_m} \circ \widetilde{\psi}_m$ for some $\phi_{j_m} \in \Phi$ and $\widetilde{\psi}_m \in \Phi^{k_m-1}$. By passing to a subsequence if necessary, we may assume that there is a $j \in \{1, \ldots, N\}$ such that $j_m = j$ for all m. Since $\{\widetilde{\psi}_m(y_m)\} \subset Y$ and Y is compact, we may pass to a subsequence again and assume that the sequence $\{\widetilde{\psi}_m(y_m)\}$ converges to some $y' \in Y$. Thus,

$$y = \lim_{m \to \infty} y_m = \lim_{m \to \infty} \phi_j(\widetilde{\psi}_m(y_m)) = \phi_j(y').$$

Therefore, $Y \subset \Phi(Y)$. $\qquad\square$

Let Y_Φ be the unique invariant compact set with respect to Φ. We will show that if μ is any invariant, Borel probability measure, its support must be Y_Φ. Therefore, without loss of generality, the search for an invariant measure may be carried out in $M_1(Y_\Phi)$.

Lemma 3.2.2. *Let μ be a Borel probability measure, invariant with respect to Φ_c. Then* $\operatorname{supp}\mu = Y_\Phi$.

Proof. Let $Y = \operatorname{supp}(\mu)$. Let U be an open set. If $U \cap Y_i \neq \emptyset$, $\phi_i^{-1}(U)$ is an open set which intersects Y. Then,

$$0 < \mu(\phi_i^{-1}(U)) = \mu_i(U),$$

implying that $\operatorname{supp}(\mu_i) \supset \overline{Y_i}$. Furthermore, if $\mu_i(U) = 0$, then $\phi_i^{-1}(U) \subset X \setminus Y$. This implies that $U \subset X \setminus \overline{Y_i}$. Thus, $\operatorname{supp}(\mu_i) = \overline{Y_i}$.

Also, for each $i \in \{1, \ldots, N\}$, $\mu_i(Y)$ has full measure by the invariance of μ. Otherwise, we would have,

$$1 = \mu(Y) = \sum_{i=1}^{N} c_i \mu_i(Y) < \sum_{i=1}^{N} c_i = 1.$$

Thus, $Y \supset \overline{Y_i}$ for $1 \leq i \leq N$, and so $\Phi(Y) \subset Y$. It follows from Lemma 3.2.1 that $Y_\Phi \subset Y$.

For the reverse inclusion, fix $R > 0$ and consider

$$B_R(Y_\Phi) = \bigcup_{y \in Y_\Phi} B_R(y) \ .$$

Let $x \in B_R(Y_\Phi)$. Choose $y \in Y_\Phi$ such that $d(x, y) \leq R$. Then, for $i = 1, 2 \ldots, N$,

$$d(\phi_i(x), \phi_i(y)) \leq r_\Phi d(x, y) \leq r_\Phi R.$$

Thus,

$$\Phi(B_R(Y_\Phi)) \subset B_{r_\Phi R}(Y_\Phi) \subset B_R(Y_\Phi),$$

and so

$$\mu(\Phi(B_R(Y_\Phi))) \leq \mu(B_{r_\Phi R}(Y_\Phi)) \leq \mu(B_R(Y_\Phi)) .$$

We have that, for $1 \leq i \leq N$, $B_R(Y_\Phi) \subset \phi_i^{-1}(\Phi(B_R(Y_\Phi)))$. Since μ is invariant,

$$\mu(\Phi(B_R(Y_\Phi))) = \sum_{i=1}^{N} c_i \mu_i(\Phi(B_R(Y_\Phi))) \geq \mu(B_R(Y_\Phi)) \sum_{i=1}^{N} c_i = \mu(B_R(Y_\Phi)).$$

Thus,

$$\mu(\Phi(B_R(Y_\Phi)) = \mu(B_{r_\Phi R}(Y_\Phi)) = \mu(B_R(Y_\Phi)).$$

Using the same argument after the j-th iteration of Φ, we obtain that for each $j \geq 1$,

$$\mu(B_{R_j}(Y_\Phi)) = \mu(B_R(Y_\Phi)), \quad R_j = (r_\Phi)^j R.$$

However, $\{B_{R_j}(Y_\Phi)\}$ is a decreasing sequence of sets whose intersection is Y_Φ. Thus,

$$\mu(Y_\Phi) = \lim_{j \to \infty} \mu(B_{R_j}(Y_\Phi)) = \mu(B_R(Y_\Phi)).$$

Furthermore, this holds for any $R > 0$. Since μ is finite, letting $R \to \infty$ gives us that $\mu(Y_\Phi) = 1$. Since Y_Φ is closed and has full measure, $Y \subset Y_\Phi$.

Therefore, $\operatorname{supp} \mu = Y_\Phi$. $\qquad\square$

For the sake of completeness, we finish the proof for the existence of a unique invariant measure.

Lemma 3.2.3. *If (X, d) is a compact metric space. Then $(M_1(X), \delta)$ is a complete metric space.*

Proof. Clearly, δ is symmetric. We show that it defines a metric on $M_1(X)$ by establishing the following 3 claims.

Claim 1: $\delta(\mu,\nu) < \infty$.

Let $\phi \in \mathrm{Lip}^1(X)$. Fix $x_0 \in X$. Let ϕ_{x_0} be the constant function which sends every point x to $\phi(x_0)$. Note that $\mu(\phi_{x_0}) = \nu(\phi_{x_0})$. Since X is compact, there is an $R > 0$ such that $X \subset B_R(x_0)$. Thus,

$$|\mu(\phi) - \nu(\phi)| \le |\mu(\phi - \phi_{x_0})| + |\nu(\phi - \phi_{x_0})|$$

$$\le \mu(d(x, x_0)) + \nu(d(x, x_0))$$

$$\le 2R.$$

Thus, $\delta(\mu,\nu) \le 2R$.

Claim 2: $\delta(\mu,\nu) = 0 \implies \mu = \nu$.

Let $\phi \in \mathrm{Lip}(X)$ with Lipschitz constant r. Then $\phi/r \in \mathrm{Lip}^1(X)$. Hence, $\mu(\phi/r) = \nu(\phi/r)$, implying $\mu(\phi) = \nu(\phi)$. By the Stone-Weierstrass Theorem, $\mathrm{Lip}(X)$ is dense in $C(X)$. Thus, $\mu(\phi) = \nu(\phi)$ for every $\phi \in C(X)$. By the Riesz Representation Theorem, $\mu = \nu$.

Claim 3: $\delta(\mu,\rho) \le \delta(\mu,\nu) + \delta(\nu,\rho)$.

We have

$$\sup_{\phi \in \mathrm{Lip}^1(X)} |\mu(\phi) - \rho(\phi)| \le \sup_{\phi \in \mathrm{Lip}^1(X)} (|\mu(\phi) - \nu(\phi)| + |\nu(\phi) - \rho(\phi)|)$$

$$\le \sup_{\phi \in \mathrm{Lip}^1(X)} |\mu(\phi) - \nu(\phi)| + \sup_{\phi \in \mathrm{Lip}^1(X)} |\nu(\phi) - \rho(\phi)|.$$

Thus, δ is a metric. To see that it is complete, let $\{\mu_n\}$ be a Cauchy sequence. Let $\phi \in \mathrm{Lip}(X)$ with Lipschitz constant r. Then $\{\mu_n(\phi/r)\}$ is a Cauchy sequence

in $\mathbb{C}$, implying that $\{\mu_n(\phi)\}$ is a Cauchy sequence as well. Again, using the fact that $\mathrm{Lip}(X)$ is dense in $C(X)$, it follows that $\{\mu_n(\phi)\}$ is a Cauchy sequence for each $\phi \in C(X)$. Since $\mathbb{C}$ is complete, $\{\mu_n(\phi)\}$ converges for each $\phi \in \mathbb{C}$.

Define

$$\mu(\phi) = \lim_{n \to \infty} \mu_n(\phi).$$

Then μ is a bounded linear functional on $C(X)$. By the Riesz Representation Theorem, $\mu \in M(X)$. Moreover, μ is clearly positive and, since the constant function $1 \in C(X)$, it is a probability measure. Therefore, $\mu \in M_1(X)$ and $\delta(\mu, \mu_n) \to 0$. $\quad\square$

Theorem 3.2.2. *Let* $\Phi = \{\phi_1, \ldots, \phi_N\}$ *be an iterated function system on a complete metric space* X. *Given a probability vector* $c \in \mathbb{R}^N$, *there is a unique* $\mu \in M_1(X)$ *invariant with respect to* Φ_c.

Proof. By Lemma 3.2.2, we may assume, without loss of generality, that $X = Y_\Phi$. That is, we may assume X is compact. We will show that Φ_c is a contraction on $(M_1(X), \delta)$. Since the metric space is complete, the result follows from the Contraction Mapping Principle.

Let $\phi \in \mathrm{Lip}^1(X)$ and $\mu, \nu \in M_1(X)$. For $1 \leq i \leq N$, the Lipschitz constant

for $\phi \circ \phi_i$ is less than or equal to r_Φ. Thus, $\phi \circ \phi_i / r_\Phi \in \mathrm{Lip}^1(X)$. Hence we have

$$
\begin{aligned}
|\Phi_c(\mu)(\phi) - \Phi_c(\nu)(\phi)| &= |\sum_{i=1}^{N} c_i(\mu(\phi \circ \phi_i) - \nu(\phi \circ \phi_i))| \\
&= |\sum_{i=1}^{N} c_i r_\Phi (\mu(\phi \circ \phi_i / r_\Phi) - \nu(\phi \circ \phi_i / r_\Phi))| \\
&\leq r_\Phi \delta(\mu, \nu) \sum_{i=1}^{N} c_i \\
&= r_\Phi \delta(\mu, \nu).
\end{aligned}
$$

Therefore,

$$
\delta(\Phi_c(\mu), \Phi_c(\nu)) \leq r_\Phi \delta(\mu, \nu).
$$

$\square$

3.2.2 Hausdorff Measures

In this section, we look at the relationship between invariant measures and Hausdorff measures. The results of this section can be found in [19], [16], and [26]. The proofs were included for the benefit of the writer.

Definition 3.2.5 (Hausdorff measure)**.** Let $p \geq 0$ and $q > 0$. For $Y \subset \mathbb{R}^d$, define

$$
H_{p,q}(Y) = \inf\{\sum_{j=1}^{\infty} (\mathrm{diam}\, E_j)^p : \ Y \subset \bigcup_{j=1}^{\infty} E_j, \ \ \mathrm{diam}\, E_j \leq q\}. \tag{3.4}
$$

Note that $H_{p,q}$ increases as q gets smaller. Hence, the limit

$$
H_p(Y) = \lim_{q \to 0} H_{p,q}(Y) \tag{3.5}
$$

exists, allowing the possibility it is infinite. For each $p \geq 0$, H_p defines a regular Borel measure on $\mathbb{R}^d$, which we call the *p-dimensional Hausdorff measure*.

Definition 3.2.6 (Hausdorff dimension)**.** The p-dimensional Hausdorff measure of any $Y \subset \mathbb{R}^d$ is equal to 0 or ∞ for most values of p. One can show that there is a unique value of p in $[0, \infty]$ where $H_p(Y)$ transitions from 0 to ∞. This value, given by

$$\dim_H(Y) = \inf\{p \geq 0 : H_p(Y) = 0\} = \sup\{p \geq 0 : H_p(Y) = \infty\}, \qquad (3.6)$$

is called the *Hausdorff dimension* of Y.

In this section, we will consider iterated function systems that consist solely of functions of the form

$$\phi : \mathbb{R}^d \to \mathbb{R}^d, \ \phi(x) = rU(x) + v \qquad (3.7)$$

where $v \in \mathbb{R}^d$, $r > 0$, and $U \in O_d(\mathbb{R})$, the group of orthogonal matrices over $\mathbb{R}^d$. Such functions are called *similitudes* and satisfy

$$\|\phi(x) - \phi(y)\| = r\|x - y\|, \qquad (3.8)$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$. In fact, (3.8) is equivalent to (3.7).

Lemma 3.2.4. *If ϕ is a similitude with Lipschitz constant $r_\phi = r$, then $\phi(H_p) = r^{-p} H_p$.*

Proof. First, note that $\|\phi(x) - \phi(y)\| = r\|x - y\|$ for all $x, y \in X$. Thus, if $E \subset X$ with $\mathrm{diam}(E) = \delta$, then $\mathrm{diam}(\phi^{-1}(E)) = r^{-1}\delta$.

Let $A \subset \bigcup_{j=1}^{\infty} E_j$ where $\mathrm{diam}(E_j) \leq \delta$ for each j. Then, $\phi^{-1}(A) \subset \bigcup_{j=1}^{\infty} \phi^{-1}(E_j)$ where $\mathrm{diam}(\phi^{-1}(E_j)) \leq \delta r^{-1}$ for each j. Thus,

$$H_{p,\delta/r}(\phi^{-1}(A)) \leq \sum_{j=1}^{\infty} \mathrm{diam}(\phi^{-1}(E_j))^p = r^{-p} \sum_{j=1}^{\infty} (\mathrm{diam}(E_j))^p.$$

Taking the infimum over all such collections $\{E_j\}$, we have

$$H_{p,\delta/r}(\phi^{-1}(A)) \leq r^{-p}H_{p,\delta}(A);$$

letting δ go to zero gives

$$H_p(\phi^{-1}(A)) \leq r^{-p}H_p(A).$$

Since similitudes are surjective, we also have that, if $\phi^{-1}(A) \subset \bigcup_{j=1}^{\infty} E_j$ with diam$(E_j) \leq \delta$ for each j, then $A \subset \bigcup_{j=1}^{\infty} \phi(E_j)$ where diam$(\phi(E_j)) \leq r\delta$. A similar argument yields

$$H_p(A) \leq r^p H_p(\phi^{-1}(A)).$$

Therefore, $H_p(\phi^{-1}(A)) = r^{-p}H_p(A)$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Definition 3.2.7 (Self-similarity). A family of similitudes $\Phi = \{\phi_1, \ldots, \phi_N\}$ is *self-similar* if $Y = Y_\Phi$ has the following properties:

- $H_p(Y) > 0$, where $p = \dim_H(Y)$, and

- for $i \neq j$, $H_p(Y_i \cap Y_j) = 0$.

A *self-similar measure* is an invariant measure with respect to a self-similar Φ.

Let $D_\Phi > 0$ be the unique value such that

$$\sum_{i=1}^{N} r_i^{D_\Phi} = 1.$$

We will see that the probability vector given by $c_i = r_i^{D_\Phi}$ is a natural choice for a self-similar iterated function system. We write the induced operator with this choice of probability vector simply as Φ.

We will show that the following condition is sufficient for Φ to be self-similar.

Definition 3.2.8 (Open set condition). Φ satisfies the *open set condition* if there is an an open set $U \subset \mathbb{R}^d$ such that $\Phi(U) \subset U$ and $U_i \cap U_j = \emptyset$ for each $i \neq j$.

Theorem 3.2.3 (Hutchinson, [26]). *Let Φ be a family of similitudes satisfying the open set condition. Then Φ is self-similar, $\dim_H(Y_\Phi) = D_\Phi$, and μ_Φ is a constant multiple of the Hausdorff measure restricted to Y_Φ.*

To prove the theorem we will need the following lemma.

Lemma 3.2.5. *Let $a, b, R > 0$. Suppose $\{U_\alpha\}$ is a collection of pairwise disjoint open sets in $\mathbb{R}^d$ such that each U_α contains a ball of radius aR and is contained in a ball of radius bR. Then, for every $x \in \mathbb{R}^d$, $B_r(x)$ intersects no more than $(1 + 2b)^d a^{-d}$ of the U_α.*

Proof. By assumption, for each α, there exist $x_\alpha, y_\alpha \in X$ such that

$$B_{ar}(x_\alpha) \subseteq U_\alpha \subseteq B_{Ar}(y_\alpha) \ .$$

Let $x_0 \in X$. If $x \in B_r(x) \cap U_\alpha$, then for each $u \in U_\alpha$,

$$\|x_0 - u_\alpha\| \leq \|x_0 - x\| + \|x - u_\alpha\| \leq r + 2Ar \ .$$

Thus, $U_\alpha \subseteq B_{r(1+2A)}(x_0)$. If $B_r(x_0)$ intersects M of the U_α, then there are M disjoint balls of radius ar inside of $B_{r(1+2A)}(x_0)$. By adding their volumes we obtain,

$$M(ar)^d \leq r^d(1 + 2A)^d.$$

Rearranging gives,

$$M \leq (1 + 2A)^d a^{-d}.$$

$\square$

Proof of Theorem 3.2.3. Let $Y = Y_\Phi$ and $D = D_\Phi$.

Claim 1: $H_D(Y) < \infty$

Let $R = \operatorname{diam}(Y)$. Then, for $\iota \in I^n(N)$,

$$\operatorname{diam}(Y_\iota) \le R(r_\phi)^n.$$

Let $p_n = R(r_\phi)^n$.

Then,

$$H_{D,p_n}(Y) \le \sum_{\iota \in I^n(N)} (\operatorname{diam}(Y_\iota))^D$$

$$= R^D \sum_{\iota \in I^n(N)} \prod_{j=1}^{n} r_{i_j}^D$$

$$= R^D (\sum_{i=1}^{N} r_i^D)^n$$

$$= R^D .$$

As $n \to \infty$, $p_n \to 0$ and so the above inequality implies that $H_D(Y) \le R^D < \infty$.

Claim 2: $H_D(Y) > 0$

To find a lower bound for $H_D(Y)$, we will use Lemma 3.2.5. Let a and A be positive numbers such that U, the open set satisfying the Definition 3.2.8, contains a ball of radius a/r_Φ and is contained in a ball of radius A. Let $C = (1 + 2A)^d a^{-d}$.

Let $\{E_j\}$ be a collection of sets covering Y, each with diameter less than or equal to 1. Let $R_j = \operatorname{diam}(E_j)$ and choose a point x_j from each set E_j. Then we have

$$Y \subset \bigcup_{j=1}^{\infty} E_j \subset \bigcup_{j=1}^{\infty} B_{R_j}(x_j) .$$

The result follows from showing that for each $x \in X$ and sufficiently small $r > 0$ we have that $\mu_\Phi(B_r(x)) \leq Cr^D$. If so, then

$$1 = \mu(Y) \leq \sum_{j=1}^{\infty} \mu(B_{R_j}(x_j)) \leq C \sum_{j=1}^{\infty} R_j^D \ .$$

Taking the infimum over all such coverings of Y gives us that

$$H_D(Y) \geq H_{D,1}(Y) \geq \frac{1}{C}.$$

Fix $1 \geq r > 0$ and let j be the unique positive integer for which $r_\Phi r \leq (r_\Phi)^j < r$. Let $I \subset \bigcup_{k=1}^{j+1} I^k(N)$ be the collection of indices satisfying $r_\Phi r \leq r_{i_1} r_{i_2} \ldots r_{i_k} < r$. Then for each $\iota \in I^{j+1}(N)$, there is a unique $\widetilde{\iota} \in I$ such that $U_\iota \subset U_{\widetilde{\iota}}$. In particular, $\widetilde{\iota} \in I^k(N)$ for some $k \in (1, \ldots, j+1)$. Thus

$$Y \subset \bigcup_{I^{j+1}(N)} \overline{U}_\iota \subset \bigcup_I \overline{U}_{\widetilde{\iota}} \ .$$

Moreover, $\{U_{\widetilde{\iota}} : \widetilde{\iota} \in I\}$ is a pairwise disjoint collection of open sets; for $\widetilde{\iota} \in I^k(N)$, $U_{\widetilde{\iota}}$ contains a ball of radius greater than ar and is contained in a ball of radius Ar, by our choice of a and A. Thus, each $U_{\widetilde{\iota}}$ contains a ball of radius ar and is contained in a ball of radius Ar. By Lemma 3.2.5, $B_r(x)$ can intersect at most C of the $U_{\widetilde{\iota}}$, and so it can intersect at most C of their closures.

Consider the iterated function system $\widetilde{\Phi} = \{\phi_{\widetilde{\iota}} : \widetilde{\iota} \in I\}$. We claim that μ_Φ is invariant with respect to $\tilde{\Phi}$.

Let $\mu = \mu_\Phi$. Since μ is invariant with respect to Φ we have

$$\mu = \sum_{\iota \in I^{j+1}(N)} \left(\prod_{k=1}^{j+1} r_{i_k} \right)^D \mu_\iota \ .$$

64

Sorting by indices in I and factoring gives

$$\mu = \sum_{\widetilde{\imath} \in I} r_{i_1}^D r_{i_2}^D \ldots r_{i_k}^D \left(\sum_{\iota \in I^{j+1-k}} r_{l_1}^D r_{l_2}^D \ldots r_{l_{j+1-k}}^D \mu_\iota \right)_{\widetilde{\imath}}$$

$$= \sum_{\widetilde{\imath} \in I} r_{i_1}^D r_{i_2}^D \ldots r_{i_k}^D \mu_{\widetilde{\imath}} \ .$$

Thus,

$$\mu(B_r(x)) = \sum_{\widetilde{\imath} \in I} r_{i_1}^D r_{i_2}^D \ldots r_{i_k}^D \mu_{\widetilde{\imath}}(B_r(x))$$

$$\leq r^D \sum_{\widetilde{\imath} \in I} \mu_{\widetilde{\imath}}(B_r(x))$$

$$\leq C r^D.$$

The last inequality holds since $B_r(x)$ intersects at most C of the $\overline{U}_{\widetilde{\imath}}$, and hence at most C of the supports for the $\mu_{\widetilde{\imath}}$.

Claim 3: The restriction of H_D to Y is invariant with respect to Φ

By Lemma 3.2.4,

$$\sum_{i=1}^N H_D(Y_i) = \sum_{i=1}^N r_i^D H_D(Y) = H_D(Y).$$

Thus, Φ is self-similar, i.e., $H_D(Y_i \cap Y_j) = 0$ for $i \neq j$.

Let $E \subset Y$. Together, self-similarity and Lemma 3.2.4 imply

$$H_D(E) = \sum_{i=1}^N H_D(E \cap Y_i)$$

$$= \sum_{i=1}^N r_i^D (H_D)_i(E).$$

$\square$

3.2.3 Cantor Measures

A *Cantor measure* is a Borel measure whose support is a Cantor set, i.e., a perfect, nowhere dense set. The class of self-similar measures provides many examples of Cantor measures. The open set condition is not quite strong enough to guarantee that the invariant measure is supported by a Cantor set. For example, consider

$$\Phi = \{\frac{1}{2}x, \frac{1}{2}(x+1)\}, \ c_1 = c_2 = \frac{1}{2}$$

on $\mathbb{R}$. Taking $U = (0,1)$, it is clear that Φ satisfies the open set condition, but μ_Φ is the restriction of Lebesgue measure to the unit interval.

However, if we assume that Φ separates its invariant set, and not just some open set, we obtain the following.

Proposition 3.2.6. *Let* $\Phi = \{\phi_1, \ldots, \phi_N\}$, $N > 1$, *be an iterated function system. Suppose that for* $Y = Y_\Phi$,

$$Y_i \cap Y_j = \emptyset, \ i \neq j.$$

Then μ_Φ *is a Cantor measure.*

Proof. Suppose Y contains an isolated point y_0. In our proof of Theorem 3.2.1, we noted that Y is the closure of the set of fixed points for functions in Φ^n for some n. Thus y_0 must be the unique fixed point for some similitude $\phi \in \Phi^n$. If y is any other point in Y, $\phi^k(y) \to y_0$ as $k \to \infty$. Hence we must have that Y consists solely of y_0. This contradicts our assumption that $N > 1$ and Φ separates Y. Therefore, Y contains no isolated points.

Moreover, the separation condition implies

$$Y = \bigcup_{\iota \in I(n)} Y_\iota$$

is a disjoint union as well. Hence, if $B_R(x)$ is contained in Y, for each $n > 0$, it must lie in Y_ι for some $\iota \in I(n)$. However,

$$\mathrm{diam}(Y_\iota) \leq (r_\Phi)^n \, \mathrm{diam}(Y) \to 0, \text{ as } n \to \infty.$$

Therefore, Y is nowhere dense. $\qquad\square$

In [31], Kahane and Salem take a different approach to constructing measures on Cantor sets. The Cantor sets themselves are built on $\mathbb{R}$ by means of repeated dissection. Associated with each resulting set Y is a continuous, monitonically increasing function $\kappa_Y : \mathbb{R} \to [0, 1]$. These κ_Y, called *Cantor-Lebesgue functions*, are fashioned so they increase only at the points of Y. Everywhere else they remain constant. Riemann-Stieltjes integration of continuous functions against κ_Y produces (by the Riesz Representation Theorem) a finite Borel measure whose support is Y.

Definition 3.2.9 (Perfect Homogeneous Sets)**.** Their construction begins with an interval $[a, b]$ of length l. Fix an integer $N \geq 2$ and select $x_1, x_2, \ldots, x_N \in [0, 1)$ such that $0 \leq x_1 < x_2 < \ldots < x_N < 1$. Let $r > 0$ satisfy $r \leq 1 - x_N$ and $r < x_n - x_{n-1}$ for $1 \leq n < N$. The first step of the dissection is to take N disjoint subintervals of length lr from $[a, b]$ of the form $[a + lx_j, a + l(x_j + r)]$. Their union is written Y_1. The same operation is then performed on the N intervals of Y_1 resulting in N^2 disjoint intervals of length lr^2 whose union we denote Y_2. Proceeding in this manner, at the

m-th step, Y_m is the disjoint union of N^m intervals of length lr^m. Explicitly,

$$Y_m = \bigcup [a + l \sum_{k=1}^{m} x_{j_k} r^{k-1}, a + l(r^m + \sum_{k=1}^{m} x_{j_k} r^{k-1})],$$

where the union is taken over all $(j_1, j_2, \ldots, j_m) \in I^m(N)$.

Finally, let Y be the intersection of the Y_m. Sets resulting from such a construction are dubbed *perfect homogeneous set of type* $(a, l, x_1, x_2, \ldots, x_N, r)$ by Kahane and Salem, referring to the fact that the same ratio of dissection, r, is used at each step.

Definition 3.2.10 (Cantor-Lebesgue Function). To construct κ_Y, define a sequence of functions $\{\kappa_m\}_{m=1}^{\infty}$ satisfying the following:

1. κ_m is 0 for $x \le a$ and 1 for $x \ge b$;

2. if the intervals of Y_m are numbered from left to right, κ_m increases linearly on the k-th interval from $(k-1)N^{-m}$ to kN^{-m};

3. κ_m is constant on the remaining intervals, taking the values needed to make it continuous.

   Note that if $m < n$,

$$\forall x \in [a, b], \ |\kappa_m(x) - \kappa_n(x)| \le \frac{1}{N^m}.$$

Hence, κ_m converges uniformly on $\mathbb{R}$. We define κ_Y to be the limit of $\{\kappa_m\}$. Then, $\kappa_Y \in C(\mathbb{R})$.

If $x \in [a, b] \setminus Y$, then $\kappa_Y(x) = \kappa_m(x)$, where m is the smallest integer such

68

that $x \notin Y_m$. Each $x \in Y$ can be written as

$$y = a + l \sum_{m=1}^{\infty} x_{j_m} r^{m-1}, \ j_m \in \{1, 2, \dots, N\}. \tag{3.9}$$

The partial sums of (3.9), which we write as y_n, gives the unique sequence of left endpoints of intervals containing x from each step. Thus,

$$\kappa_Y(y) = \lim_{n \to \infty} \kappa_m(y_n) = \sum_{m=1}^{\infty} (j_m - 1) N^{-m}.$$

Riemann-Stieltjes integration of continuous functions against κ_Y defines a positive linear functional on $C([a, b])$. By the Riesz Representation Theorem, there is a unique Borel measure μ_{κ_Y} with support in $[a, b]$ such that for each $f \in C([a, b])$,

$$\int_a^b f d\kappa_Y = \int f d\mu_{\kappa_Y}.$$

Notice that the Riemann-Stieltjes integral vanishes for any $f \in C([a, b])$ whose support lies in an open interval disjoint from Y since κ_Y is constant on such intervals. On the other hand, for a nonnegative $f \in C([a, b])$ whose support lies in an open interval intersecting Y, we must have that $\int_a^b f d\kappa_Y > 0$. This implies that $\text{supp } \mu_{\kappa_Y}$ is precisely Y, and hence, is a Cantor measure. We will call this class of Cantor measures, *perfect homogeneous measures.*

Moreover, since the integration by parts formula holds for Riemann-Stieltjes integration, it follows that the perfect homogeneous measures can also be thought of as the distributional derivatives of the Cantor-Lebesgue functions.

The Fourier-Stieltjes transform of μ_{κ_Y} is given by

$$\widehat{\mu_{\kappa_Y}}(\gamma) = \int e^{-2\pi i \gamma x} d\mu_{\kappa_Y} = \int_a^b e^{-2\pi i \gamma x} d\kappa_Y.$$

Using the fact that κ_Y increases by N^{-m} over successive left endpoints of Y_m, we obtain the following approximation for $\widehat{\mu_Y}$:

$$\widehat{\mu_{\kappa_Y}}(\gamma) \sim N^{-m} e^{-2\pi i \gamma a} \sum \prod_{k=0}^{m-1} e^{-2\pi i \gamma l x_{j_k} r^{m-1}},$$

where the sum is taken over all N^m indices $(j_1, j_2, \ldots, j_m) \in I^m(N)$. Rewriting this as the product of sums gives

$$\widehat{\mu_{\kappa_Y}}(\gamma) \sim e^{-2\pi i \gamma a} \prod_{k=0}^{m-1} (N^{-1} \sum_{j=1}^{N} e^{-2\pi i \gamma l x_j r^k}).$$

Letting $m \to \infty$, we have

$$\widehat{\mu_{\kappa_Y}}(\gamma) = e^{-2\pi i \gamma a} \prod_{k=0}^{\infty} (N^{-1} \sum_{j=1}^{N} e^{-2\pi i \gamma l x_j r^k}).$$

The product converges pointwise since the modulus of each term is clearly less than or equal to 1.

We construct an iterated function system Φ whose invariant measure has the same Fourier-Stieltjes transform as μ_{κ_Y}. By the uniqueness of the transform, the two measures will be the same.

With the construction of Y in mind, define the following family of similitudes. For $1 \le j \le N$, let

$$\phi_j(x) = r(x - a) + a + l x_j, \quad \Phi = \{\phi_1, \ldots, \phi_N\}.$$

Note that

$$Y_m = \bigcup_{\iota \in I^m(N)} \phi_\iota([a, b]).$$

Moreover, letting $\mu = \mu_\Phi$, iteratively using the invariance of μ gives

$$\widehat{\mu}(\gamma) = \mu(e^{-2\pi i \gamma x})$$

$$= \widehat{\mu}(r\gamma)e^{-2\pi i \gamma (1-r)a}(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j})$$

$$= \widehat{\mu}(r^2\gamma)e^{-2\pi i \gamma (r-r^2)a}e^{-2\pi i \gamma (1-r)a}(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j r})(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j})$$

$$= \widehat{\mu}(r^2\gamma)e^{-2\pi i \gamma (1-r^2)a}(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j r})(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j})$$

$$= \quad \vdots$$

$$= \widehat{\mu}(r^m\gamma)e^{-2\pi i \gamma (1-r^m)a}\prod_{k=0}^{m-1}(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j r^k})$$

Since $r < 1$ and μ has total mass 1, $\lim_{m\to\infty}\widehat{\mu}(r^m\gamma) = \widehat{\mu}(0) = 1$. Letting $m \to \infty$

above then yields

$$\widehat{\mu}(\gamma) = e^{-2\pi i \gamma a}\prod_{k=0}^{\infty}(N^{-1}\sum_{j=1}^{N}e^{-2\pi i \gamma l x_j r^k}) \ .$$

Thus we have proved the following.

Theorem 3.2.4. *Let μ_{κ_Y} be the Cantor measure associated with the perfect homo-*
geneous set of type $(a, l, x_1, x_2, \ldots, x_N, r)$. Then μ_{κ_Y} is the unique Borel probability
measure invariant under the iterated function system $\Phi = \{\phi_1, \ldots, \phi_N\}$, where for
$1 \le j \le N$, $\phi_j(x) = r(x - a) + a + l x_j$. $\square$

3.2.4 Haar Measures on Cantor Sets

In this section, we consider only iterated function systems $\Phi = \{\phi_1, \ldots, \phi_N\}$
which separate their invariant set, as in Proposition 3.2.6. We will show that a

group structure can be induced on Y_Φ such that the invariant measure associated with the equal weight probability vector is a Haar measure.

Let Q_N be the space $\{0, \ldots, N-1\}^{\mathbb{N}}$ endowed with the product topology. By Tychonoff's theorem, this topology makes Q_N a compact space. Also, note that each $q \in Q_N$ has a neighborhood basis consisting of *simple cylinders*, i.e. sets of the form

$$C_m(q) = \{u \in Q_N : u_i = q_i, \ 1 \le i \le m\}.$$

Lemma 3.2.7. *Every open set $U \subset Q_N$ can be written as the pairwise disjoint, countable union of simple cylinders.*

Proof. For each $q \in U$, let $C(q)$ be the largest simple cylinder around q contained in U. Then

$$U = \bigcup_{q \in U} C(q).$$

Since Q_N is second countable, this cover of U has a countable subcover. Furthermore, it is clear from the definition of simple cylinders that $C(q) \cap C(q') \ne \emptyset$ if and only if one is contained in the other. Thus, by passing to a second subcover if necessary, there exists a sequence $\{q_k : k \in \mathbb{N}\} \subset U$ such that

$$U = \bigcup_{k=1}^{\infty} C(q_k)$$

and the union is pairwise disjoint. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Lemma 3.2.8. *Y_Φ is homeomorphic to Q_N.*

Proof. Let $Y = Y_\Phi$. Since Φ separates Y, for each $y \in Y$, there is a unique sequence

of indices $\iota_k \in I^k(N)$ such that $y \in Y_{\iota_k}$. Hence, the function

$$\pi : Q_N \to Y, \ q \mapsto \bigcap_{k=1}^{\infty} Y_{q_k \ldots q_1}$$

provides a bijection between Q_N and Y.

The assumption that Φ separates Y also implies that $\{Y_{q_k \ldots q_1} : k \in \mathbb{N}\}$ is a neighborhood basis for $\pi(q)$. Thus, to show that π is continuous, it suffices to show that each $\pi^{-1}(Y_{q_k \ldots q_1})$ is open. But,

$$\pi^{-1}(Y_{q_k \ldots q_1}) = C_k(q).$$

A continuous bijection from a compact set to a Hausdorff space is a homeomorphism. Therefore, π is a homeomorphism. $\qquad\square$

Addition on Q_N is defined inductively. For $q_1, q_2 \in Q_N$,

$$(q_1 + q_2)(1) = (q_1(1) + q_2(1))(\mathrm{mod}\ N)$$

$$(q_1 + q_2)(k) = (q_1(k) + q_2(k))(\mathrm{mod}\ N) + \left\lfloor \frac{q_1(k-1) + q_2(k-1)}{N} \right\rfloor, \ k > 1.$$

where $\lfloor \cdot \rfloor$ rounds down to the nearest integer. Plainly stated, the group structure on Q_N is given by long addition. Working from left to right, we add entrywise, carrying the one where necessary.

With this definition of addition, Q_N is a compact, abelian topological group. Namely, it is $\mathbb{Z}_N$, the N-adic integers.

Theorem 3.2.5. *Let ν be the Haar measure on $\mathbb{Z}_N$, normalized so that $\nu(\mathbb{Z}_N) = 1$. Then $\mu_\Phi = \pi(\nu)$.*

Proof. Let $\mu = \mu_\Phi$ and $Y = Y_\Phi$. Since π is a homeomorphism, it suffices to show that the pullback of μ, $\pi^{-1}(\mu)$, is translation invariant on $\mathbb{Z}_N$. From the preceding lemmas, it follows that it is sufficient to check translation invariance for simple cylinders.

Fix $u \in Q_N$ and $m \in \mathbb{N}$. Then

$$\forall q \in Q_N, \ q + C_m(u) = C_m(q+u).$$

However, for every simple cylinder of the form $C_m(q)$, $\pi(C_m(q)) = Y_{q(1)...q(m)}$. The Φ invariance of μ implies that

$$\mu(Y_{q(1)...q(m)}) = \frac{1}{N^m} \sum_{\iota \in I^m(N)} \mu_\iota(Y_{q(1)...q(m)}).$$

We assumed that Φ separates Y. Hence, for every $m \in \mathbb{N}$,

$$Y = \bigcup_{\iota \in I^m(N)} Y_\iota$$

is a pairwise disjoint union. This implies $\mu_\iota(Y_{q(1)...q(m)}) \neq 0$ if and only if $\iota = (q(1), \ldots, q(m))$. Furthermore, $\mu_\iota(Y_\iota) = \mu(Y) = 1$. Therfore, it follows that

$$\forall q \in Q_N, \ \pi^{-1}(\mu)(C_m(q)) = \mu(Y_{q(1)...q(m)}) = \frac{1}{N^m}.$$

$\square$

Since $\mathbb{Z}_N$ is a compact abelian group, its dual group, $\widehat{\mathbb{Z}_N}$, is an orthonormal basis for $L^2(\mathbb{Z}^N)$. The previous theorem allows us to push this orthonormal basis forward into $L^2(\mu_\Phi)$, giving us the following useful corollary.

Corollary 3.2.6. *Let* $\{\epsilon_p : \ p \in \widehat{\mathbb{Z}_N}\}$ *be the characters for* $\mathbb{Z}_N$. *Then* $\{\epsilon_p \circ \pi^{-1}\}$ *is an orthonormal basis for* $L^2(\mu_\Phi)$. $\square$

3.3 Spectral Properties of the Middle-Third Cantor Measure

As mentioned before, in general, singular measures provide little to work with, turning the most basic questions from the classical setting of PW_Ω into very difficult problems. To gain any insight into the problem, it is useful to work with measures that provide enough structure for one to get their hands dirty with computations. The class of self-similar measures provides such a structure.

The interest in the spectral properties of self-similar measure began with a surprising result from Jorgensen and Pedersen.

Theorem 3.3.1 (Jorgensen and Pedersen, [30]). *Let* $\Phi = \{\frac{1}{4}x, \frac{1}{4}(x+2)\}$. *The set* $S \subset \mathbb{R}$ *consisting of all finite sums of the form*

$$\sum_{k=0}^{N-1} c_k 4^k, \ c_k \in \{0, 1\}$$

generates a Fourier basis for $L^2(\mu_\Phi)$.

Numerous papers have subsequently been written on this family of measures, though, all have fallen short when it comes to understanding the spectral properties of the middle-third Cantor measure. We too will come up short in this section. However, we will provide some first steps, including an outline for constructing Bessel spectra.

3.3.1 Basic Facts

For the remainder of the section, unless stated otherwise, μ will be the middle-third Cantor measure and $\Omega = \text{supp}\,\mu$ will be the middle-third Cantor set. We begin

by establishing some well known properties of μ.

First, we note that, from our work in Section 3.2, there are three different ways of thinking of μ:

- As the invariant measure for $\Phi = \{\frac{1}{3}\gamma, \ \frac{1}{3}(\gamma + 2)\}$.

- As the distributional derivative of $\kappa_{1/3}$, the Cantor-Lebesgue function for the middle-third Cantor set.

- As the pushforward of Haar measure on the 2-adic integers.

These are certainly not the only ways of thinking about μ. For example, μ is also a member of the class of Bernoulli convolutions, for which their is an extensive body of research.

We can also compute the Fourier-Stieltjes transform of μ using our computations in Section 3.2.3:

The method of repeated dissection provides the familiar construction of Ω:

$$\Omega_0 = [0, 1]$$

$$\Omega_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$$

$$\vdots$$

$$\Omega_k = \bigcup_{j=0}^{2^k - 1} \Omega_j^k, \ \text{where } \Omega_j^k = [\omega_j, \omega_j + 3^{-k}] \text{ and } \omega_j = 2 \sum_{n=1}^{l} c_j(n) 3^{-j}, \ c_j(n) \in \{0, 1\}$$

$$\Omega = \bigcap_{k=0}^{\infty} \Omega_k$$

Thus,

$$\widehat{\mu}(x) = \prod_{k=1}^{\infty} \left(\frac{1}{2} \left(1 + e^{-2\pi i x \frac{2}{3^k}} \right) \right)$$

$$= e^{-2\pi i x (\sum_{k=1}^{\infty} 3^{-k})} \prod_{k=1}^{\infty} \cos\left(\frac{2\pi i x}{3^k} \right)$$

$$= e^{-\pi i x} \prod_{k=1}^{\infty} \cos\left(\frac{2\pi i x}{3^k} \right) \qquad (3.10)$$

Proposition 3.3.1. $\widehat{\mu}$ *is an entire function of type* 2π.

Proof. For $z \in \mathbb{C}$ we have,

$$|\widehat{\mu}(z)| = |\int e^{-2\pi i z \omega} d\mu(\omega)|$$

$$\leq \int e^{2\pi |z| \omega} d\mu(\omega)$$

$$\leq e^{2\pi |z|}$$

since $\Omega \subset [0,1]$.

Also, if $\mathfrak{c}$ is any closed curve in $\mathbb{C}$, we have

$$\int_{\mathfrak{c}} \int e^{-2\pi i z \omega} d\mu(\omega) dz = \int \int_{\mathfrak{c}} e^{-2\pi i z \omega} dz d\mu(\omega) = 0.$$

Hence by Morera's theorem $\widehat{\mu}$ is entire. $\qquad \square$

Definition 3.3.1. The set

$$Z(\widehat{\mu}) = \{z \in \mathbb{C} : \widehat{\mu}(z) = 0\}$$

is the *zero set* for $\widehat{\mu}$.

The zero set is an important object of study for both entire functions and spectral measures. For instance, we have the following simple necessary condition for any S to be a Fourier spectrum for a given measure μ.

Proposition 3.3.2. *Let S be a Fourier spectrum for μ. Then*

$$S - S \subset Z(\widehat{\mu}).$$

Proof. This is simply a restatement of the requirement that S generate an orthogonal set in $L^2(\mu)$:

$$\widehat{\nu}(s' - s) = \langle e_s, e_{s'}\rangle_\mu = 0.$$

$\square$

By computing the zero set for the middle-third μ, this simple criterion will show that it cannot be a Fourier measure.

Lemma 3.3.3.

$$Z(\widehat{\mu}) = \{\frac{(2k + 1)3^k}{4} : k \in \mathbb{Z}, j \in \mathbb{N}\}$$

.

Proof. The set of zeros described is precisely the set of all zeros of functions of the form $\cos(\frac{2\pi z}{3^k})$. Hence, the result will follow from showing that if $\widehat{\mu}(z) = 0$, then there is an N such that

$$\prod_{k=1}^{N} \cos(\frac{2\pi z}{3^k}) = 0. \tag{3.11}$$

Since $\widehat{\mu}(0) = 1$, there is a neighborhood, U, of the origin in $\mathbb{C}$ such that $|\widehat{\mu}(z)| \geq c > 0$ for all points $z \in U$. For any $z \in \mathbb{C}$, there is an $N > 0$ such that $z3^{-N} \in U$ and so $\widehat{\mu}(z3^{-N}) \geq c$. If (3.11) never vanishes for z,

$$\widehat{\mu}(z) = e^{c_0 \Im(z)} \left(\prod_{j=1}^{N} \left| \cos\left(\frac{2\pi z}{3^j}\right) \right| \right) \widehat{\mu}(\frac{z}{3^N}),$$

where $c_0 \in \mathbb{R}$ is the appropriate constant. Hence, if (3.11) does not equal 0, $\widehat{\mu}(z) > 0$.

$\square$

Proposition 3.3.4. *Any set $S \subset \mathbb{R}$ such that $E(S)$ consists of mutually orthogonal functions in $L^2(\mu)$ contains at most 2 elements.*

Proof. Suppose S contains 3 distinct members. Then there exist $k, n \in \mathbb{Z}$ and $m \geq j \in \mathbb{N}$ such that

$$s_1 - s_2 = \frac{(2k+1)3^j}{4}, \text{ and } s_2 - s_3 = \frac{(2n+1)3^m}{4}.$$

Then

$$s_1 - s_3 = (s_1 - s_2) + (s_2 - s_3) = (2k + 1 + (2n+1)3^{m-j})\frac{3^j}{4}.$$

But

$$(2k + 2n3^{m-j} + 3^{m-j} + 1) \in 2\mathbb{Z}.$$

This contradicts Lemma 3.3.3. □

3.3.2 Constructing Spectra

Not until recently was any real progress made in understanding the spectral properties of μ. In [15], it was shown that Bessel spectra exist. While their proof provides significant insight into the necessary structure of Bessel spectra, it not quite constructive. Our main goal in this section is to present a simple avenue through which construction might be achieved.

First, we make the simple observation that, when constructing Bessel spectra, one must avoid what we call *lacunary harmonics*.

Definition 3.3.2. For each $(x, n) \in \mathbb{R} \times \mathbb{Z}$, define

$$L(x, n) = \{x_j = x + n3^j : j \in \mathbb{Z}\}.$$

79

Proposition 3.3.5. *If S is a Bessel spectra for $L^2(\mu)$, then, for each $(x, n) \in \mathbb{R} \times \mathbb{Z}$,*

$$\text{card}(S \cap L(x, n)) < \infty.$$

Proof. Let J be the set of indices for the $x_j \in L(x, n) \cap S$. Since S is a Bessel spectrum for μ, there is a $B > 0$ such that $\forall x \in \mathbb{R}$,

$$\sum_{s \in S} |\widehat{\mu}(s - x)|^2 \leq B.$$

Therefore,

$$\sum_{j \in J} \prod_{k=1}^{\infty} \cos^2\left(\frac{2\pi(x_j - x)}{3^k}\right) = \sum_{j \in J} \prod_{k=1}^{\infty} \cos^2\left(\frac{2\pi n}{3^k}\right) \leq B.$$

Thus,

$$\text{card}(L(x, n) \cap S) = \text{card}(J) \leq \frac{B}{|\widehat{\mu}(n)|^2} < \infty.$$

$\square$

Our approach is motivated by the various constructions of the Cantor set and its measure. In each case, the object was attained as the limit of a sequence of operations performed on some known quantity. A natural question is to ask if spectra can be constructed in a similar manner. We call this the *limiting approach* and establish its viability for constructing Bessel spectra in the following theorem.

Theorem 3.3.2. *Let $\{\mu_k : k \in \mathbb{N}\}$ be a sequence of probability measures with support in $[-R, R]$ for some $1 \leq M < \infty$. Suppose that the sequence converges to μ in the weak-$*$ topology. For each k, let*

$$S_k = \{s_j^k : j \in I_k \subset \mathbb{Z}\}$$

be a Bessel spectrum for μ_k with bound B_k. Assume that the sequences have been indexed in such a way that $I_k \subseteq I_{k+1}$. Let $I = \cup I_k$. Suppose that

    i. for each $j \in I$, $\lim\limits_{k \to \infty} s_j^k$ exists, and

    ii. $\liminf\limits_{k \to \infty} B_k < \infty$.

Let

$$s_j = \lim_{k \to \infty} s_j^k.$$

Then $S = \{s_j\}_{j \in I}$ is a Bessel spectrum for μ.

First, we need the following lemma. It is a specific case of a well known fact about frames in Hilbert spaces [9].

Lemma 3.3.6. *To show that S is a frame spectrum for $L^2(\mu)$, it suffices to show that the frame inequality holds for a dense set of functions.*

Proof. Let $\mathcal{C}$ be a dense subset of $L^2(\mu)$ and let $L_S : \mathcal{C} \to l^2(S)$ be the linear operator defined by $L_S(\phi) = \{\langle \phi, e_s \rangle_\mu : s \in S\}$. The upper frame bound gives us that

$$\|L_S(\phi)\|_2^2 \leq B\|\phi\|_\mu^2, \forall \phi \in \mathcal{C}.$$

Hence, L_S is a bounded linear operator on $\mathcal{C}$. Since $\mathcal{C}$ is dense in $L^2(\mu)$, L can be uniquely extended to a bounded linear operator defined on all of $L^2(\mu)$.

Fix $\phi \in L^2(\mu)$. Choose a sequence $\{\phi_n\} \subset \mathcal{C}$ converging to ϕ. Then,

$$\sum_{s \in S} |\langle \phi, e_s \rangle_\mu|^2 \leq \liminf_{n \to \infty} \|L_S(\phi_n)\|_2^2$$

$$\leq \liminf_{n \to \infty} B\|\phi_n\|_\mu^2$$

$$= B\|\phi\|_\mu^2.$$

Thus,

$$L_S(\phi) = \{\langle \phi, e_s \rangle_\mu : s \in S\}.$$

Moreover, by the continuity of L_S,

$$A\|\phi\|_\mu^2 = A \lim_{n \to \infty} \|\phi_n\|_\mu^2$$

$$\leq \lim_{n \to \infty} \|L_S(\phi_n)\|_2^2$$

$$= \|L_S(\phi)\|_2^2.$$

Therefore, the frame inequality is satisfied for every $\phi \in L^2(\mu)$. $\qquad\square$

Proof of Theorem 3.3.2. Fix $\phi \in C([-R, R])$. Let $L_{S_k}(\phi) : I \to \mathbb{C}$ be the sequence defined by

$$L_{S_k}(\phi)(j) = \begin{cases} \langle \phi, e_{s_j^k} \rangle_{\mu_k} & \text{if } j \in I_k \\ 0 & \text{if } j \notin I_k \end{cases}$$

Let $j \in I$. For sufficiently large k,

$$|\langle \phi, e_{s_j^k} \rangle_{\mu_k} - \langle \phi, e_{s_j} \rangle_\mu| \leq |\langle \phi, e_{s_j^k} - e_{s_j} \rangle_{\mu_k}| + |\langle \phi, e_{s_j} \rangle_{\mu_k} - \langle \phi, e_{s_j} \rangle_\mu|.$$

Let $M = \|\phi\|_\infty$. Then,

$$|\langle \phi, e_{s_j^k} - e_{s_j} \rangle_{\mu_k}| \leq M \max_{[-R,R]} |e_{s_j^k}(x) - e_{s_j}(x)|$$

$$\leq 2\pi M R |s_j^k - s_j| \to 0.$$

Additionally,

$$|\langle \phi, e_{s_j} \rangle_{\mu_k} - \langle \phi, e_{s_j} \rangle_\mu| \to 0$$

since $\mu_k \xrightarrow{w^*} \mu$. Thus, $L_{S_k}\phi \to L_S(\phi)$ pointwise on I.

Hence,

$$\|L_S(\phi)\|_2^2 \le \liminf_{k \to \infty} \|L_{S_k}(\phi)\|_2^2 \le \liminf_{k \to \infty} B_k \|\phi\|_{\mu_k}^2 = B\|f\|_\mu^2.$$

Therefore, by Lemma 3.3.6 and the fact that $C([-R,R])$ is dense in $L^2(\mu)$, S is a Bessel spectra for μ. $\qquad\square$

Normalizing the sets in the construction of Ω provides a sequence of measures in $M_1(\mathbb{R})$ useful for computing integrals in $L^2(\mu)$. For $k \in \mathbb{N}_0$, let ν_k be the restriction of Lebesgue measure to Ω_k and let $\mu_k = \left(\frac{3}{2}\right)^k \nu_k$.

Proposition 3.3.7. *The sequence of measures $\{\mu_k\}$ converge to μ in the weak-$*$ topology.*

Proof. Consider a step function of the form

$$\phi(x) = \sum_{j=1}^{2^N} c_j \mathbb{1}_{\Omega_j^N}(x), \quad c_j \in \mathbb{C}.$$

For $k \ge N$, $\mu_k(\Omega_j^N) = \dfrac{1}{2^N} = \mu(\Omega_j^N)$. Thus, for $k \ge N$,

$$\mu_k(\phi) = \sum_{j=1}^{2^N} c_j \mu_k(\Omega_j^N) = \sum_{j=1}^{2^N} c_j \mu(\Omega_j^N) = \mu(\phi).$$

Let $\psi \in C([0,1])$ and define

$$\psi_N(x) = \sum_{j=1}^{2^N} \psi(a_j) \mathbb{1}_{K_j^N}(x).$$

Since ψ is uniformly continuous on $[0,1]$, we can choose N large enough so that $|\psi(x) - \psi_N(x)| < \epsilon$ on Ω_N. Thus, for $k \ge N$,

$$|\mu_k(\psi) - \mu(\psi)| \le |\mu_k(\psi) - \mu_k(\psi_N)| + |\mu_k(\psi_N) - \mu(\psi_N)| + |\mu(\psi_N) - \mu(\psi)|$$

$$= |\mu_k(\psi - \psi_N)| + |\mu(\psi - \psi_N)| < 2\epsilon.$$

Therefore, $\mu_k(\psi) \to \mu(\psi)$. $\qquad\square$

The middle-third IFS provides many examples of sequences of measures converging to μ in the weak-$*$ topology. For any probability measure ν with support in $[0,1]$, we have $\Phi^k(\nu) \xrightarrow{w^*} \mu$. In fact, Proposition 3.3.7 can be reformulated in this way, since $\Phi^k(\mu_0) = \mu_k$.

We use Proposition 3.3.7 for the following computation.

Proposition 3.3.8. *Let* $\chi(\gamma) = \sum_{k=1}^{2^m} c_k \mathbb{1}_{\Omega_k^m}(\gamma)$. *Then*

$$|\langle \chi, e_x \rangle_\mu| = \frac{1}{2^m} |\widehat{\mu}(\frac{x}{3^m})| |\sum_{k=1}^{2^m} c_k \overline{e_x}(\omega_k)| \tag{3.12}$$

and

$$\|\chi\|_\mu^2 = \frac{1}{2^m} \sum_{k=1}^{2^m} |c_k|^2. \tag{3.13}$$

Proof. (3.13) is straightforward since

$$\|\chi\|_\mu^2 = \int |\chi(\gamma)|^2 d\mu = \sum_{k=1}^{2^m} |c_k|^2 \mu(\Omega_k^m) = \frac{1}{2^m} \sum_{k=1}^{2^m} |c_k|^2.$$

Define μ_n as in Proposition 3.3.7. For $j \in \mathbb{N}$, let $n = m + j$. Then

$$|\langle \chi, e_x \rangle_{\mu_n}| = \left| \left(\frac{3}{2}\right)^n \sum_{k=1}^{2^m} c_k \sum_{l=1}^{2^j} \int_{\omega_l^k}^{\omega_l^k + 3^{-n}} \overline{e_x}(\gamma) d\gamma \right|$$

where the ω_l^k are the 2^j left endpoints of the $\Omega_l^n \subset \Omega_k^m$. Letting ω_k be the left endpoint of Ω_k^m we have that

$$\omega_l^k = \omega_k + 2 \sum_{i=1}^{j} \frac{c_l(i)}{3^{m+i}}, \quad c_l \in \{0,1\}^j$$

84

Thus,

$$|\langle \chi, e_x \rangle_{\mu_n}| = \left| \left(\frac{3}{2} \right)^n \sum_{k=1}^{2^m} c_k \sum_{l=1}^{2^j} \frac{\overline{e_x}(\omega_l^k + 3^{-n}) - \overline{e_x}(\omega_l^k)}{2\pi x} \right|$$

$$= \left| \frac{3^n}{\pi x} \sin\left(\frac{\pi x}{3^n} \right) \right| \left| \frac{1}{2^n} \sum_{k=1}^{2^m} c_k \sum_{l=1}^{2^j} \overline{e_x}(\omega_l^k) \right|$$

$$= \left| \frac{3^n}{\pi x} \sin\left(\frac{\pi x}{3^n} \right) \right| \left| \frac{1}{2^m} \sum_{k=1}^{2^m} c_k \overline{e_x}(\omega_k) \right| \left| \frac{1}{2^j} \sum_{l=1}^{2^j} \prod_{i=1}^{j} \overline{e_x}\left(\frac{2c_l(i)}{3^{m+i}} \right) \right|$$

$$= \left| \frac{3^n}{\pi x} \sin\left(\frac{\pi x}{3^n} \right) \right| \left| \frac{1}{2^m} \sum_{k=1}^{2^m} c_k \overline{e_x}(\omega_k) \right| \left| \prod_{i=1}^{j} \frac{1 + \overline{e_x}\left(\frac{2}{3^{m+i}} \right)}{2} \right|$$

$$= \left| \frac{3^n}{\pi x} \sin\left(\frac{\pi x}{3^n} \right) \right| \left| \frac{1}{2^m} \sum_{k=1}^{2^m} c_k \overline{e_x}(\omega_k) \right| \left| \prod_{i=1}^{j} \cos\left(\frac{2\pi x}{3^{m+i}} \right) \right|$$

Taking the limit as $j \to \infty$, we have

$$|\langle \chi, e_x \rangle_\mu| = \left| \frac{1}{2^m} \sum_{k=1}^{2^m} c_k \overline{e_x}(\omega_k) \right| \left| \prod_{i=1}^{\infty} \cos\left(\frac{2\pi x}{3^{m+i}} \right) \right|.$$

$\square$

The class of simple functions given in Proposition 3.3.8 provides a useful tool for the analysis of μ. Because of the totally disconnected structure of μ, this class of functions is dense in all of $L^2(\mu)$. Hence, they provide a nice computational step towards constructing Bessel and frame sequences in $L^2(\mu)$.

Note that by taking a finite set $S_N \subset S$, we have

$$\sum_{n=1}^{N} |\langle \chi, e_{s_n} \rangle_\mu|^2 = \sum_{n=1}^{N} \left| \frac{1}{2^m} \sum_{k=1}^{2^m} c_k \overline{e_{s_n}}(\omega_k) \right|^2 \left| \prod_{i=1}^{\infty} \cos\left(\frac{2\pi s_n}{3^{m+i}} \right) \right|^2$$

$$\leq \frac{1}{4^m} \sum_{n=1}^{N} \left| \sum_{k=1}^{2^m} c_k \overline{e_{s_n}}(\omega_k) \right|^2$$

$$= \frac{\|Tc\|^2}{4^m},$$

85

where

$$T = (\overline{e_{s_n}}(\omega_k)) \in M_{N,2^m}(\mathbb{C}). \tag{3.14}$$

Thus, the problem of constructing Bessel spectra can be reduced to the problem of constructing a family of matrices of type (3.14) in such a way that the matrix norms can be controlled.

The following proposition shows that simply requiring T to be injective is enough to yield Fourier frames in $L^2(\mu_m)$.

Proposition 3.3.9. *Let $N \geq 2^m$ and choose $\{s_n\}_{n=1}^N \subset \mathbb{R}$ such that*

$$T = \left(\overline{e_{s_n}}(\omega_k)\right) \in M_{N,2^m}(\mathbb{C})$$

is 1:1. Then

$$S = \bigcup_{n=1}^{N} \{s_n + 3^m \mathbb{Z}\}$$

is a frame spectrum for $L^2(\mu_m)$.

Proof. Since T is 1:1, there exist $A, \ B > 0$ such that

$$\forall v \in \mathbb{C}^{2^l}, \quad A\|v\|^2 \leq \|Tv\|^2 \leq B\|v\|^2.$$

Fix $\phi \in L^2(\mu_l)$. Let $\phi_j(x) = \phi(x + a_j)$ and $v_x = \langle \phi_j(x) \rangle_{j=1}^{2^m}$. For each n,

$E(s_n + 3^m\mathbb{Z})$ is an orthonormal basis for $L^2([0, 3^{-m}])$. Thus,

$$\sum_{n=1}^{N}\sum_{j\in\mathbb{Z}}|\langle\phi, e_{s_n+j3^m}\rangle_{\mu_m}|^2 = \sum_{n=1}^{N}\sum_{j\in\mathbb{Z}}\left|\left(\frac{3}{2}\right)^m\sum_{l=1}^{2^m}\int_{\omega_l}^{\omega_l+3^{-m}}\phi(\omega)\overline{e_{s_n+j3^m}}(\omega)d\omega\right|^2$$

$$= \sum_{n=1}^{N}\sum_{j\in\mathbb{Z}}|3^m\int_{0}^{3^{-m}}\left(2^{-m}\sum_{l=1}^{2^m}\phi_l(\omega)\overline{e_{s_n}}(\omega_l)\right)\overline{e_{s_n+j3^m}}(\omega)d\omega|^2$$

$$= \sum_{n=1}^{N}\left(\frac{3}{4}\right)^m\int_{0}^{3^{-m}}|\sum_{l=1}^{2^m}\phi_l(\omega)\overline{e_{s_n}}(\omega_l)|^2d\omega$$

$$= \left(\frac{3}{4}\right)^m\int_{0}^{3^{-m}}\|Tv_x\|^2dx$$

$$\leq B\left(\frac{3}{4}\right)^m\int_{0}^{3^{-m}}\|v_x\|^2dx$$

$$= \frac{B}{2^m}\|\phi\|_{\mu_m}$$

The lower frame bound follows similarly. □

Using more technical machinery, it has been shown that when $N = 2^m$ the same condition implies that S will generate a Riesz basis when Ω is the finite disjoint union of equal length intervals, spaced regularly in $\mathbb{R}$ [7].

Finally, in [14], the authors showed that the geometry of self-similar measures manifests itself in the spectral properties in an intriguing manner. They showed that the Hausdorff dimension of the support of the measure will place certain restrictions on the dimension of the spectra.

Definition 3.3.3. The *upper and lower α-Beurling density* of a discrete $S \subset \mathbb{R}$ are

given by

$$D_\alpha^+(S) = \limsup_{r\to\infty} \sup_{x\in\mathbb{R}} \frac{\mathrm{card}(S\cap Q_r(x))}{r^\alpha} \qquad (3.15)$$

$$D_\alpha^-(S) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}} \frac{\mathrm{card}(S\cap Q_r(x))}{r^\alpha} \qquad (3.16)$$

The *upper and lower Beurling dimension* of S is defined by

$$\mathrm{Dim}^+(S) = \sup\{\alpha > 0 : D_\alpha^+ > 0\}$$

$$\mathrm{Dim}^-(S) = \inf\{\alpha < \infty : D_\alpha^- < \infty\}$$

The notion of Beurling dimension was introduced in [10] where they used it to classify families of Gabor pseudoframes.

In [14], it was shown that the Beurling dimension of a Bessel spectrum cannot exceed the Hausdorff dimension of supp μ. They also showed that if a frame spectrum satisfies a certain technical condition, its Beurling dimension must equal the Hausdorff dimension.

We provide a simple proof of the result for Bessel spectra of the middle-third Cantor measure using our computation in Lemma 3.3.8.

Theorem 3.3.3. *Let S be a Bessel spectrum for μ and let $\alpha = \log_3 2$. Then*

$$D_\alpha^+(S) < \infty. \qquad (3.17)$$

Proof. Suppose $D_\alpha^+(S) = \infty$. Then given $N > 0$, there is an $r_0 > 0$ such that

$$\forall R \geq r_0, \ \exists x_0 \in \mathbb{R} : \ \mathrm{card}(S\cap B_R(x)) \geq NR^\alpha. \qquad (3.18)$$

Since $\widehat{\mu}(0) = 1$, given $C \in (0,1)$, there is an $R_0 > 0$ such that $\widehat{\mu}(x) \geq C$ for every $x \in B_{R_0}(0)$.

Choose $m \in \mathbb{N}$ such that $3^m R_0 \geq r_0$. Let $x_0 \in \mathbb{R}$ satisfy (3.18) for $R = 3^m R_0$.

Let $\chi(\gamma) = e_{x_0}(\gamma) \mathbb{1}_{[0,3^{-m}]}(\gamma)$. If $B > 0$ is the Bessel bound for S, then

$$\frac{B}{2^m} = B\|\chi\|_\mu^2 \geq \sum_{s \in S} |\langle \chi, e_s \rangle_\mu|^2$$

$$= \sum_{s \in S} \frac{1}{4^m} |\widehat{\mu}(\frac{s - x_0}{3^m})|^2$$

$$\geq \frac{1}{4^m} \sum_{s \in B_R(x_0)} |\widehat{\mu}(\frac{s - x_0}{3^m})|^2$$

$$\geq \frac{C^2}{4^m} \operatorname{card}(S \cap B_R(x_0))$$

$$\geq \frac{NC^2}{4^m}(3^m R_0)^\alpha = \frac{NC^2}{2^m} R_0^\alpha.$$

Thus,

$$B \geq NC^2 R_0^\alpha.$$

Since N can be arbitrarily large, this contradicts the assumption that S is a Bessel spectrum. $\qquad \square$

Part II

Laplacian Eigenmaps

Chapter 4

Laplacian Eigenmaps and Image Analysis

4.1 Introduction

Technological advancements over the last several decades have greatly increased our ability to collect, transmit, and store large amounts of information. For example, in a 2010 article, the Washington Post estimated that the NSA siphons intelligence from 1.7 billion emails, calls, etc., intercepted *daily* [47]. Such advancements in data acquisition require the innovation of efficient means of gleaning, representing, and analyzing the information buried therein. Many areas of research, dimension reduction theory, data mining, machine learning, pattern recognition, feature extraction, neural networks, and compressive sensing to name a few, have emerged as a result.

Laplacian Eigenmaps (LE) is a member of a large catalogue of techniques, called *kernel methods*, that lie in the intersection of dimension reduction theory and feature extraction. Broadly stated, the objective of a kernel method is to find a mapping, called a *kernel*, from a data set X into a feature space Y. The kernel is typically constructed so that the dimensions of the feature space correspond to properties of interest in the data set, yielding a better representation of the data.

The primordial example of a kernel method is Principal Component Analysis (PCA).

Example 4.1.1 (PCA). Let $X = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^D$. Define

$$\widehat{x_n} = x_n - c, \ \widehat{X} = [\widehat{x}_1 \ \widehat{x}_2 \ \ldots \ \widehat{x}_N]$$

where

$$c = \frac{1}{N} \sum_{n=1}^{N} x_n$$

is the centroid for X. Form the $D \times D$ covariance matrix

$$\Phi = \widehat{X}\widehat{X}^T.$$

Diagonalize Φ and sort the eigenvectors in decreasing order according to their eigenvalues. The kernel $\phi_d : \mathbb{R}^D \to \mathbb{R}^d$ is the matrix whose rows correspond to the first d eigenvectors of Φ.

4.1.1   LE Algorithm

In contrast with PCA, LE is a *non-linear* kernel method. The algorithm was conceived and developed theoretically by Belkin and Nyogi [1] and has become popular tool for a variety of dimension reduction problems and clustering problems.

Laplacian Eigenmaps belongs to a family of non-linear kernel methods motivated by the problem of finding low-dimensional manifold embeddings. We assume that $X \subset \mathbb{R}^D$ lies on a manifold in $\mathbb{R}^D$ which preserves the local geometric structure of the data set. The goal is to find a kernel $\phi : X \to \mathbb{R}^d$, which reflects aspects of the local and global geometric structure of X.

The algorithm is comprised of three steps.

(1) Construct a k-nearest neighbor (kNN-)graph.

For each $x_n \in X \subset \mathbb{R}^D$, draw an undirected edge between x_n and its k-nearest neighbors in the data set with respect to the Euclidean metric on $\mathbb{R}^D$. The resulting graph is represented by its $N \times N$ *adjancency matrix* G:

$$G(i,j) = \begin{cases} 1 & \text{if there is an edge connecting } x_i, x_j \\ 0 & \text{otherwise} \end{cases}$$

Note that the k-nearest neighbor relation is not symmetric, however, by making the edges undirected, the graph, and hence G, will be. Imposing symmetry has several advantages. Symmetric matrices are easier to handle computationally. Moreover, directed graphs are more likely to have connectivity issues, presenting a barrier to acquiring any information about the global geometric structure of X.

(2) Weight the graph.

A *weight matrix* for G is a nonnegative, symmetric, $N \times N$ matrix W with the same sparsity structure as G. The *degree matrix* corresponding to W is the diagonal matrix D, with diagonal entries given by

$$D(i) = \sum_{j=1}^{N} W(i,j).$$

The *weighted graph Laplacian* is given by

$$L = D - W.$$

Since W is symmetric, L is symmetric. The Laplacian is also positive semidefinite,

since

$$y^T L y = y^T D y - y^T W y$$

$$= \frac{1}{2} \left(\sum_{i=1}^{N} y_i^2 D(i) + \sum_{j=1}^{N} y_j^2 D(j) - 2 \sum_{i,j=1}^{N} y_i y_j W(i,j) \right)$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} (y_i^2 + y_j^2 - 2 y_i y_j) W(i,j)$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} (y_i - y_j)^2 W(i,j) \geq 0 \tag{4.1}$$

(3) Solve the eigenproblem

The d coordinate mappings for the LE embedding $\phi : X \to \mathbb{R}^d$ are given by the eigenvectors solving the generalized eigenproblem

$$L y = \lambda D y \tag{4.2}$$

for the d smallest non-zero eigenvalues. These solutions can be found by iteratively solving

$$\underset{\substack{\langle y, D y \rangle = 1 \\ y \in Y_0^\perp}}{\arg \min} \, y^T L y \tag{4.3}$$

where $Y_0^\perp$ is the orthogonal complement of the subspace spanned by the previous solutions and the kernel of quadratic form. It follows from the computation in (4.1) that the minimizers of $y^T L y$ produce an embedding that will preserve the local geometry of X. That is, ϕ will map close points in X to close points in $\mathbb{R}^d$.

Also, note that $\dim(\ker(\Phi)) \geq 1$, since $\Phi(y) = 0$ if y is constant. If G is connected, any two vertices x_i, x_j can be connected by a path in G. Hence there is a sequence of indices $i = i_0, i_1, \ldots, i_n = j$ such that $W(i_k, i_{k-1}) > 0$ for $1 \leq k \leq n$. Thus, if $\Phi(y) = 0$, it must be constant along the path connecting x_i, x_j, implying it

94

must be constant everywhere. Thus, if G is connected, $\dim(\ker(\Phi)) = 1$. A similar argument shows that, in general, $\dim(\ker(\Phi))$ is equal to the number of connected components of G. If we assume G to be connected, the zero solution provides no further information and hence can justifiably be ignored.

4.1.2 Approximate k-Nearest Neighbors

An exact neighborhood construction can take a long time for large data sets. For a set of N data points in $\mathbb{R}^D$ requires $O(DN^2)$ time. To reduce the runtime, we implemented the approximate nearest neighbor search developed in [8]. Their method employs a Divide and Conquer routine via *spectral bisection*:

Divide:

The data is recursively divided into two overlapping sets. A parameter $0 < \alpha < 1$ is controls the amount of overlap. At each step the largest singular value and its corresponding vectors are computed using the Lanczos algorithm. The points are separated by their relation to the hyperplane determined by the left singular vector.

Brute:

The data set is divided into smaller and smaller pieces until they reach some predetermined manageable size. Then exact kNN is performed on each piece.

Conquer:

For points that lie in the overlap, the k-nearest neighbors are selected from search results in both pieces.

The amount of overlap determines the complexity of the the algorithm:

$$O(DN^t), \quad t = \frac{1}{1 - \log_2(1 + \alpha)}.$$

In our experience, the method worked very well, reducing computation time from minutes to seconds, without sacrificing accuracy.

4.1.3 Geometric Motivation

The algorithm is motivated in large part by Spectral Geometry. In particular, the typical choice of weights is motivated by the *heat kernel*:

$$x, y \in \mathbb{R}^D, t \in \mathbb{R} \quad h_t(x, y) = (4\pi t)^{-\frac{D}{2}} e^{-\frac{\|x-y\|^2}{4t}}$$

which appears when solving the *heat equation*:

$$u_t + \Delta u = 0$$

$$u(x, 0) = f(x)$$

With this in mind, the nonzero entries of the weight matrix are chosen to be

$$W(i, j) = \left(e^{-\frac{\|x_i - x_j\|^2}{4\sigma^2}} \right). \tag{4.4}$$

The analogy was made concrete theoretically by Belkin and Niyogi [2]. They proved that, assuming the data set X_t is an increasingly fine, uniformly distributed sample of a compact manifold $M \subset \mathbb{R}^D$, the system converges in probability to Δ_M, the Laplace-Beltrami operator acting on $C^\infty(M)$,

- pointwise – that is, the system converges weakly as operators acting on the space of smooth functions – and,

- spectrally – that is, the eigenvalues and eigenfunctions converge as well.

4.2 Applications

In this section, we provide two examples of applications of LE. The first is a novel use of LE for dimension *expansion*; the dataset is embedded into a space of greater dimension than the starting space. This highlights the fact that LE is primarily a tool for feature extraction and can be used for finding deterministic clusterings of a dataset. The second is an example of its utility as a tool for meaningful dimension reduction, and will provide a basis for comparison for the results in the next chapter.

4.2.1 Analysis of SAS Images

This project was born out of a collaboration with a group from the Naval Surface Warfare Center (NSWC), Panama City. The collaboration was made possible thanks to a grant from the Office of Naval Research (ONR). The project was motivated by the following problem:

Is it possible to distinguish the contents of a barrel from sonar data?

Two classes of barrels were submerged, distinguished by their contents. One class contained sand and the other contained an unknown synthetic material called filler E (fE). Synthetic aperture sonar (SAS) data was collected for each submerged barrel at a high and low frequency. The team from NSWC had developed a beam-forming technique in which they solved an inverse Radon transform problem and

created an image resembling a topographical map of the ocean floor from the raw SAS data.
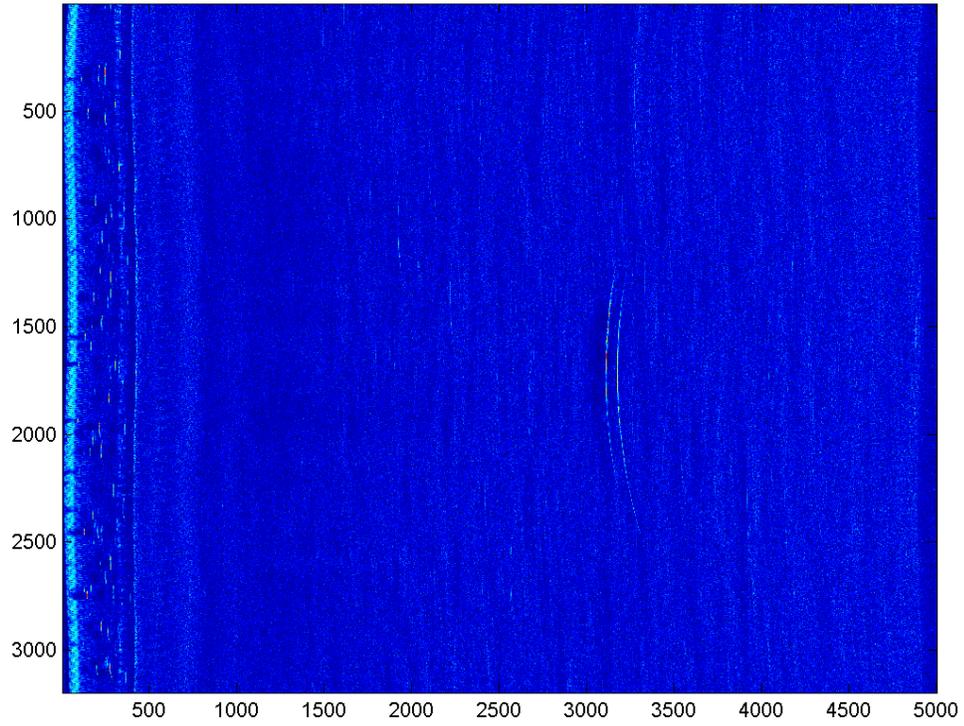
Figure 4.1: Raw SAS data
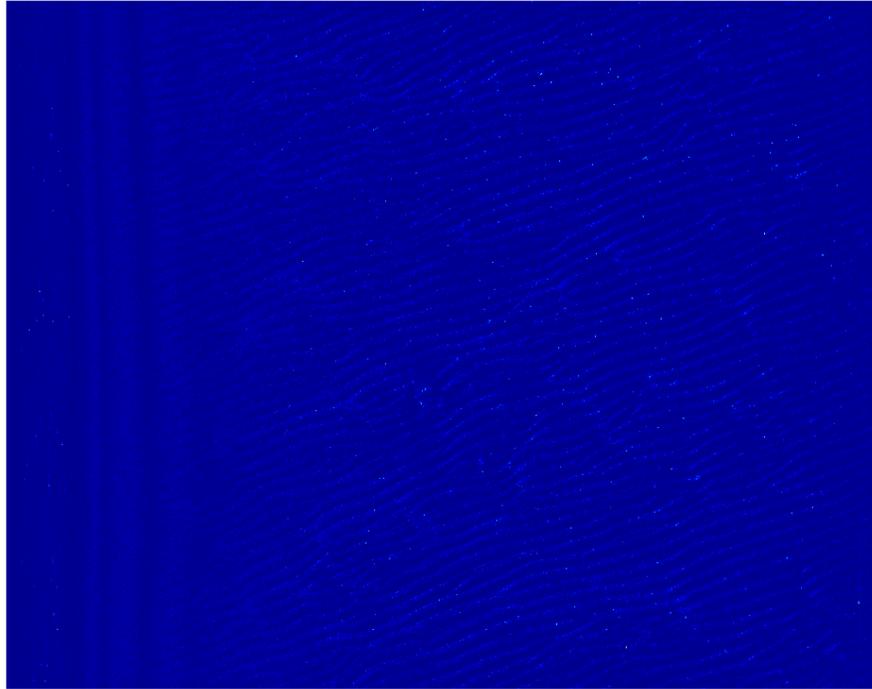
Figure 4.2: Beamformed image

Our task was to see if the high and low frequency beamformed images provided

enough variation in the dataset to develop a classification routine.

Visually, the two frequencies of any given barrel are very similar. This makes

it difficult to detect any variation between the two classes of barrels.

Using a dimension reduction technique on a two dimensional dataset seems

somewhat contradictory. However, the number of eigenvectors determined by LE is

limited by the number of data points in the set, not its dimensionality. In this case,

we were limited by the $200 \times 100 = 20,000$ pixels in each cropped image. Moreover,

the eigenvectors essentially provide a representation of the clustering done in the

99

construction of the nearest neighbor graph. Hence, it is reasonable to hypothesize that if sufficient variation exists to distinguish the two classes of barrels, LE will provide evidence of this.

Our results were encouraging. First, they provided a visually diverse array of images for each barrel (Figure 4.3). Moreover, we noticed a pattern in the mappings of the two different classes. The energy in the images of each sand barrel tend to be concentrated on the edges of the barrel, whereas for the fE barrels, at least one dimension of the mapping indicates a significant amount of energy located in the interior of the barrel as well.
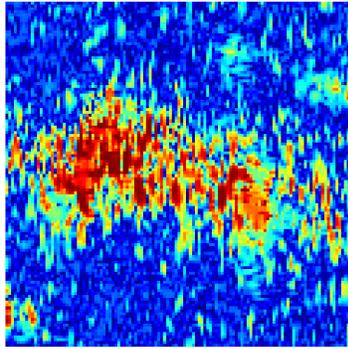
k=1
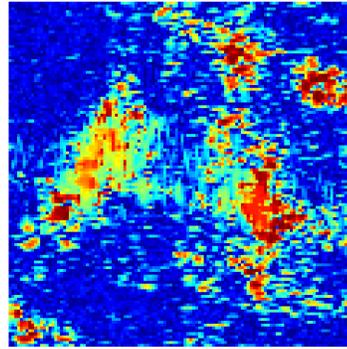
k=2
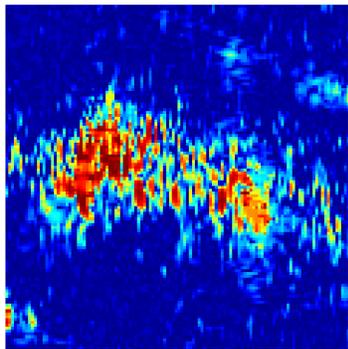
k=3
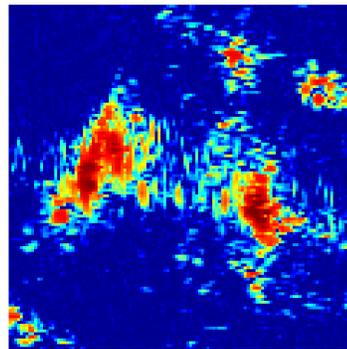
k=4
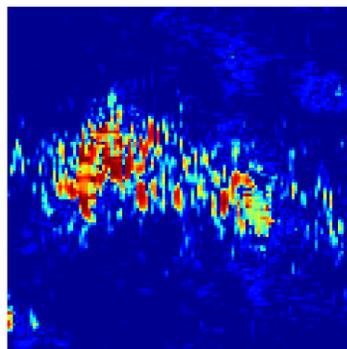
k=5

Figure 4.3: Eigenbarrel components for a fE barrel.

Using this information, we designed the following classification routine. For illustration, two barrels were chosen from each class, B_S and B_{fE}.

(1)) For each barrel, find the convex hull of points clearly lying on its edges.
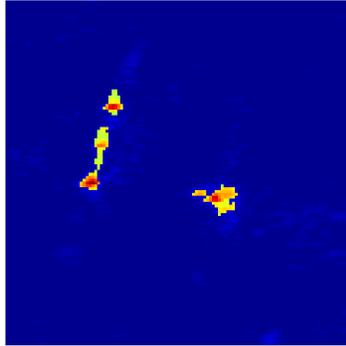
Edges for B_S
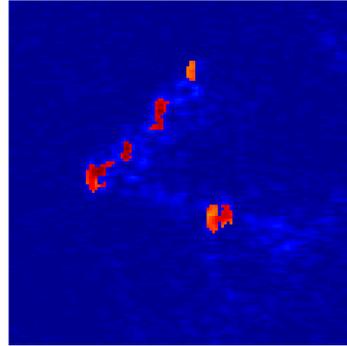
Edges for B_{fE}.
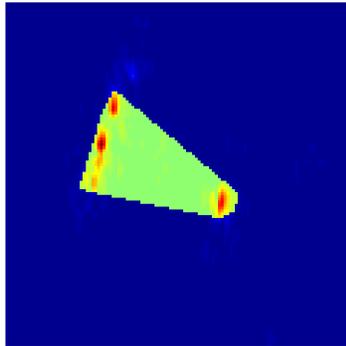
Convex hull for B_S.
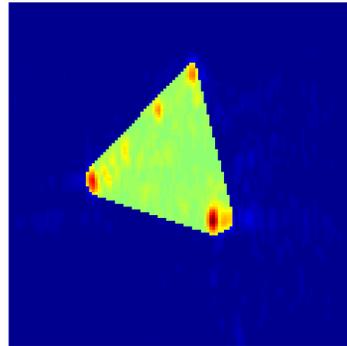
Convex hull for B_{fE}.

Figure 4.4: Definition of the convex hull for a barrel.

The edges of each barrel are identified via thresholding in the low frequency image. The magnitude of the points on the edges clearly dominate the rest of the image. The shape and size of each hull varies significantly in accordance with the manner in which each barrel was laying on the ocean floor. However, each

hull captures the region of the image we are chiefly interested in, namely the area immediately following the visible edges of the barrel.

(2) Perform a histogram analysis of the exterior and interior regions of the hull.
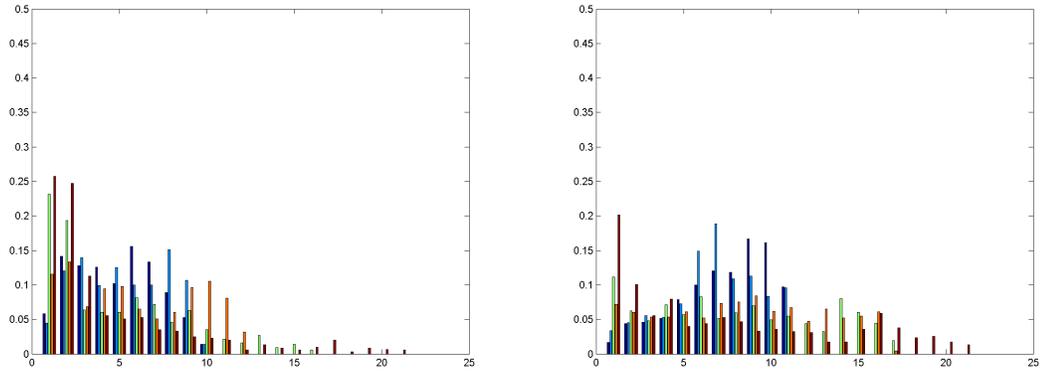
Interior histogram for B_S

Interior histogram for B_{fE}.
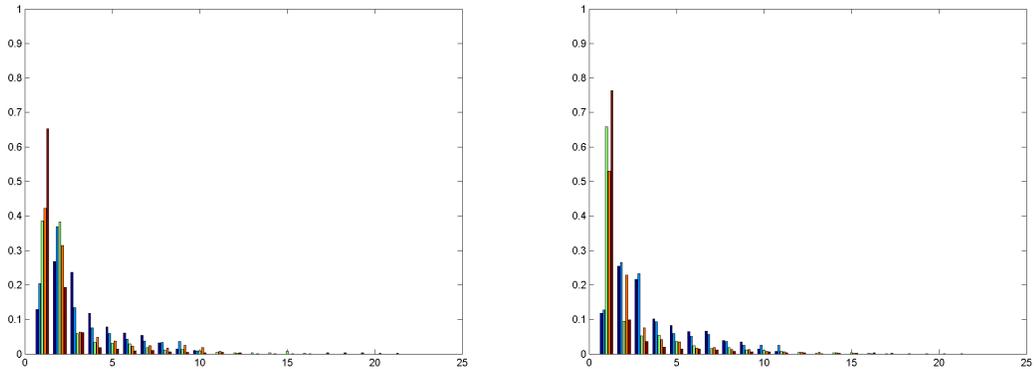
Exterior histogram for B_S.

Exterior histogram for B_{fE}.

Figure 4.5: Interior and exterior histograms.

The interior of the hull is defined to be hull minus the edge points. The exterior is the complement of the hull in the image. A histogram is created for the exterior and interior regions in each of the 5 components of the eigenbarrel. The components

of each eigenbarrel are normalized so that $\|\phi_k\|_\infty = 1$. The unit interval is divided into 20 bins of equal size and the resulting histogram is divided by the total number of pixels in the corresponding region so that we end up with a representation of the proportional distribution of pixels in each bin. In Figure 4.5, larger bin numbers correspond to larger magnitudes.

The exterior histogram is computed to indicate the level of noise present in the area surrounding each barrel. This is important because these artifacts will occasionally appear in the components of the mapped image and overlap the hull of the eigenbarrel. Figure 4.6 provides an example of this phenomenon.
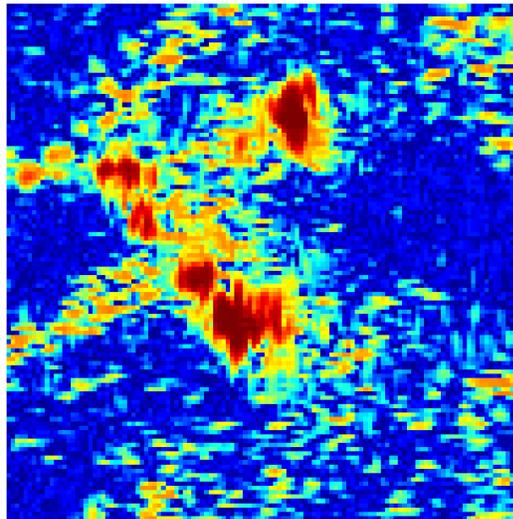
Figure 4.6: A barrel in a noisy environment.

Thus, we can create a relative histogram for each image by subtracting the exterior from the interior. In Figure 4.7, the visual differences we originally noted are visible in the tails of the histograms.

Relative histogram for B_S Relative histogram for B_{fE}.
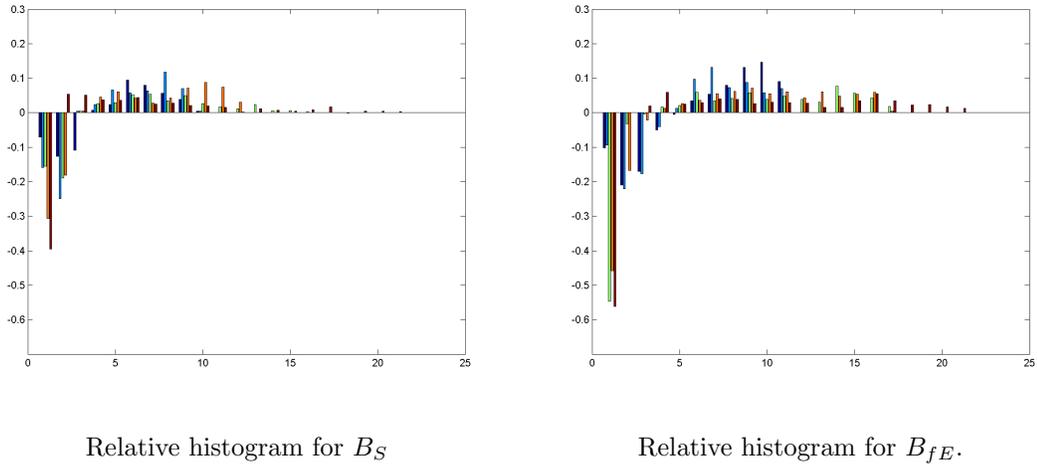
Figure 4.7: Relative histograms.

(3) Classify using SVM.

The relative histogram is the feature vector that we use for classification. A support vector machine with Gaussian kernel was used to classify the set of feature vectors. In total, there were 61 barrels—32 fE and 29 sand. We used 17 of these for training—8 sand and 9 fE. Our results are given in Figure 4.8. Performance on each class and overall accuracy are provided.
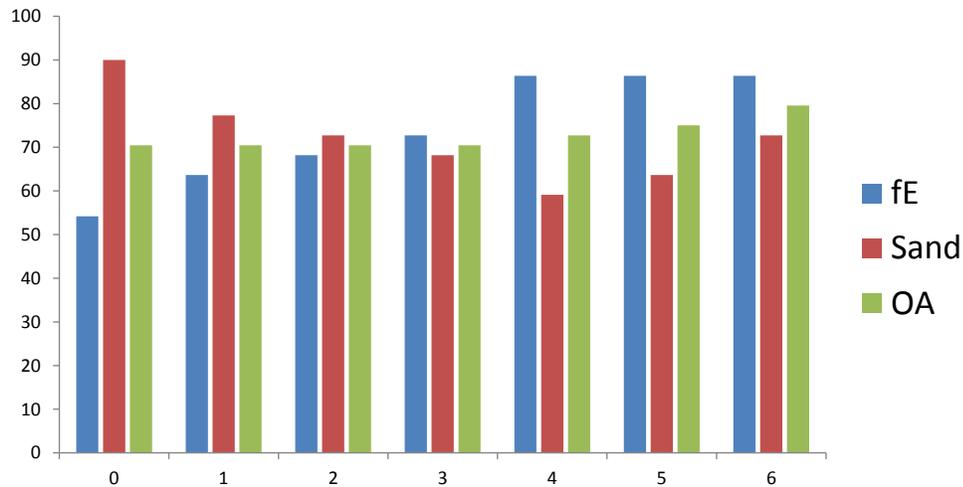
Figure 4.8: Classification results for LE on SAS data.

The bottom axis indicates the number of dimensions used in the mapping. The 0-dimensional mapping is the histogram analysis done on the original data set. Overall accuracy increased from 70% to 80%. This provides some evidence of the usefulness of LE as a feature extraction tool, even in a low dimensional setting.

4.2.2 Analysis of HSI

A typical camera takes pictures using three bands of light–red, blue, and green. Hyperspectral images (HSI) are captured in 100-200 narrow bands of light. The narrow bands provide a finer scale of classification. A fine spectrum is especially

useful when the picture is being taken over long distances, e.g. from a plane or satellite. Hence, HSI technology is found in a diverse array of applications requiring remote sensing, such as searching for mineral deposits, studying forest health, and intelligence gathering.

The large number of bands also makes it infeasible for analysts to process the image manually, however. HSI analysis is an area of research devoted to designing supervised, semi-supervised, and unsupervised methods for detecting and classifying objects of interest within hyperspectral images.

HSI datasets X are typically represented as an $N \times M \times D$ datacube.
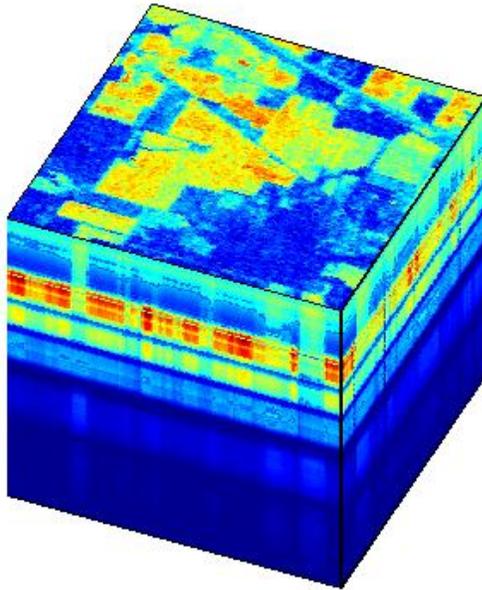
Figure 4.9: Indian Pines hyperspectral cube

Each $x \in X$ is the spectral vector for a pixel $(n, m) \in \{1, \ldots N\} \times \{1, \ldots, M\}$, which corresponds to a physical location in the image.

We introduce two datasets, Pavia University (Figure 4.10) and Indian Pines (Figure 4.11), that will be used in the next chapter. Our analysis here will serve as a baseline for comparison. Both datasets are freely available on the web.

False color image   Ground truth
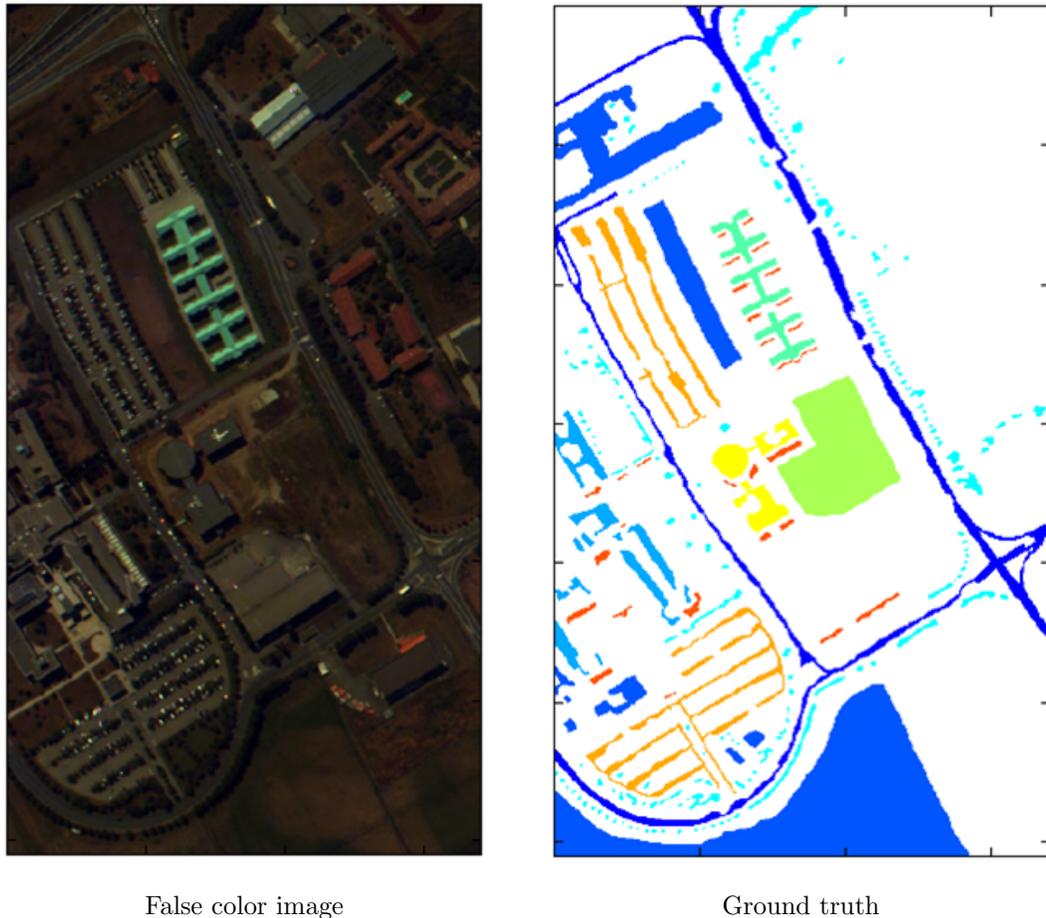
Figure 4.10: Pavia University dataset

The Pavia University image is a 610×340 pixel image. It contains 115 spectral bands with approximately 1.3 meter resolution. It was acquired in a flyover of Pavia, Italy using a ROSIS sensor. There is a collection of 42776 ground truth pixels associated with this data set, covering a total of 9 classes. We chose this data set for two reasons. First, its large size would present a suitable computational

challenge for our methods. Second, the ground truth had a particular structure. In Chapter 5, we will use spatial/spectral techniques to analyze the image. Many datasets have ground truth that are spatially biased, e.g. Indian Pines. The ground truth for Pavia University, however, is fairly spread out and intermingled, providing a sufficient challenge for our methods.

False color image   Ground truth
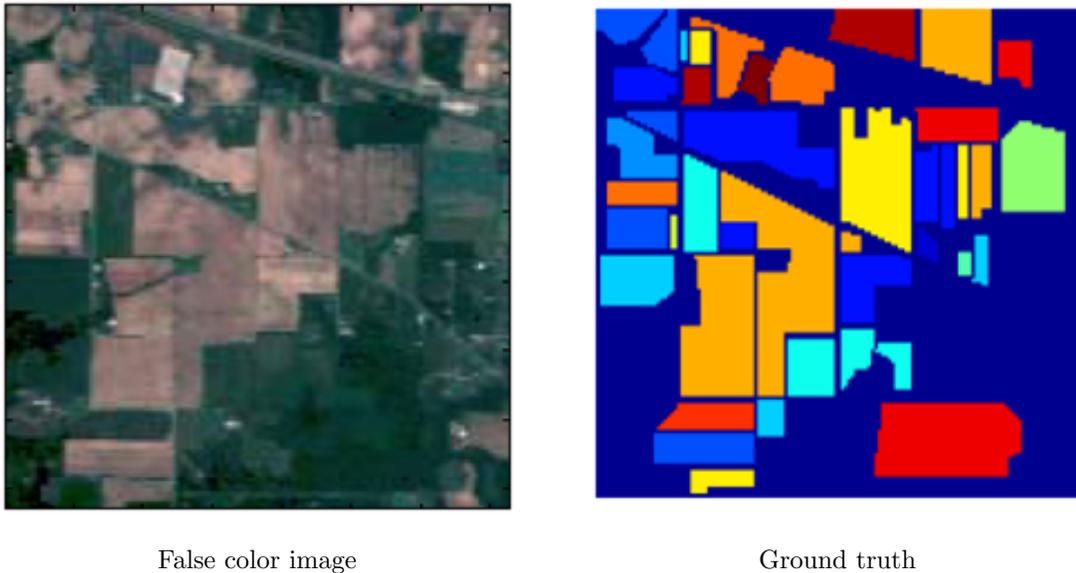
Figure 4.11: Indian Pines dataset

The Indian Pines image is a 145×145 pixel image that contains 224 spectral bands with approximately 20 meter resolution; 4 bad bands were removed, leaving a $145 \times 145 \times 220$ datacube. It was acquired in a flyover of northwest Indiana using an AVIRIS spectrometer. There is a collection of 10249 ground truth pixels associated with the dataset covering a total of 16 classes. Indian Pines was selected primarily because, as we will see, LE performs rather poorly on it. We will use Indian Pines as a test for the methods we develop in the next chapter based on Schroedinger

Eigenmaps.

Our implementation of LE for these datasets proceeds as follows:

(1) Compute the approximate kNN-graph

Store the graph as a sparse $N \times N$ matrix, where N is the number of pixels. For both Indian Pines and Pavia University, the number of nearest neighbors was fixed at 20. We did not analyze the optimal number of nearest neighbors for either dataset. We simply chose a number that provided a connected graph.

(2) Form the Laplacian for the graph.

Store the Laplacian as a sparse $N \times N$ matrix. The separation parameter, σ, was fixed at 1.

(3) Solve the generalized eigenproblem

Ignoring the trivial solution, find the first d solutions to the generalized eigenproblem. For Pavia University, we set $d = 25$ and used the JDQZ eigensolver built in to the Dimension Reduction Toolbox based on the QZ algorithm. For Indian Pines, we set $d = 50$ and use the built in MATLAB sparse eigensolver, EIGS.

(4) Classify

Use the embedding as our feature vector for each pixel. Classification was done using *vector angle analysis*: A certain percentage of ground truth from each class are randomly selected for training. For each pixel, the vector angle between its feature vector and those of the training is computed. The point is then mapped to the class which minimizes the vector angle. This can be repeated n-times and the classification is determined by the mode.

We chose vector angle classification because of its simplicity. More sophisticated algorithms would produce better results, however we are interested in testing how well LE, and our extension of LE, extract features from the hyperspectral image. For Pavia University we selected 1% of the ground truth pixels for each class for training. For Indian Pines, we selected 10% .

Results

In Figures 4.12 and 4.13 we provide some selected eigenvectors for our datasets.

k=1   k=2   k=3
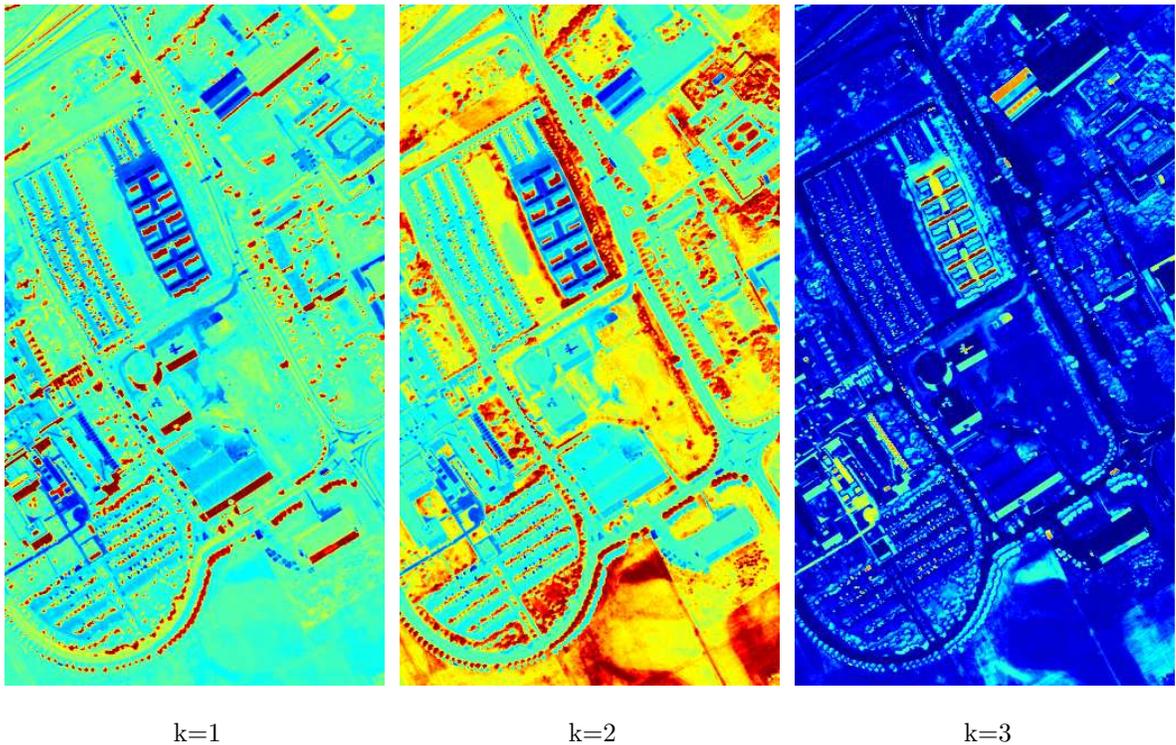
Figure 4.12: Selected eigenvectors from the Pavia University embedding

Note that several features are brought out in the first three dimensions. In $k = 1$, the shadows of the buildings stand out, in $k = 2$, the meadows are apparent, and in $k = 3$, several structures are highlighted.

k=1   k=2

k=11
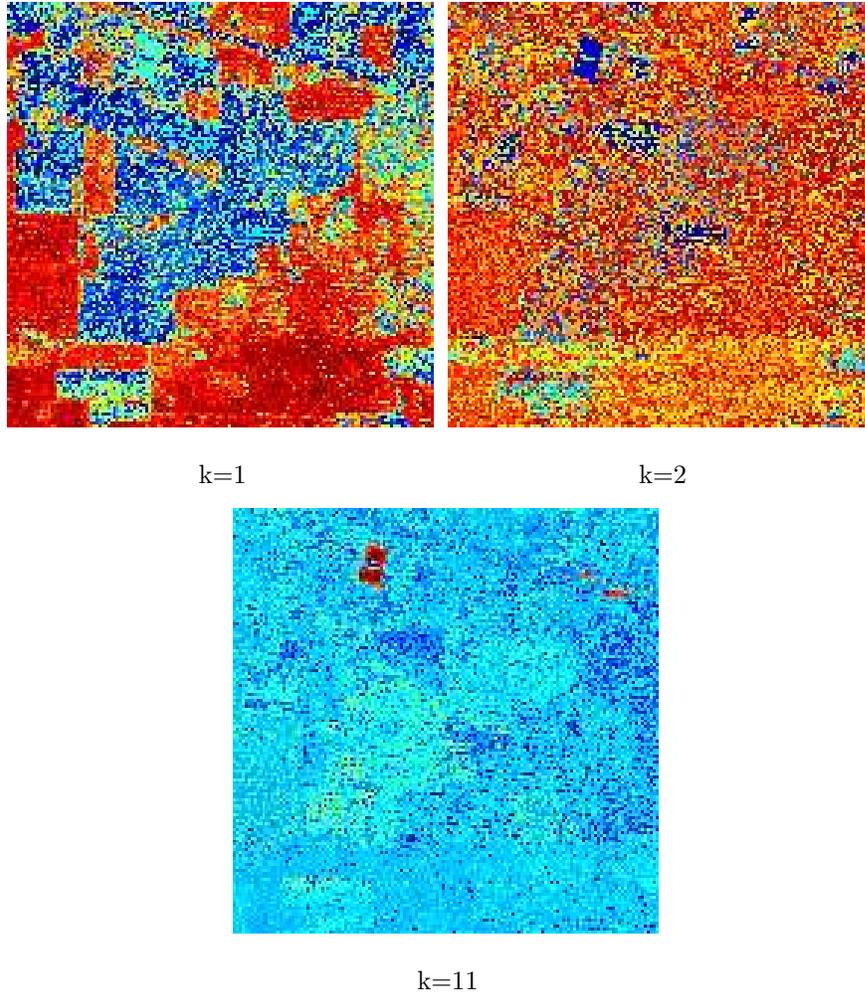
Figure 4.13: Selected eigenvectors from the Indian Pines embedding

In contrast, the Indian Pines embedding is much worse. The images are very grainy, making it hard to distinguish anything. One of the reasons why is that the image is fairly homogeneous. The picture is of flat farmland, and so most of the objects in the image have a very similar spectral signature.

The class maps and overall accuracies are provided in Figures 4.14 and 4.15.
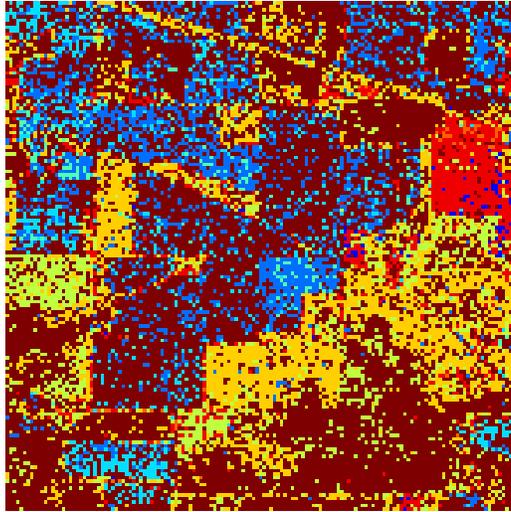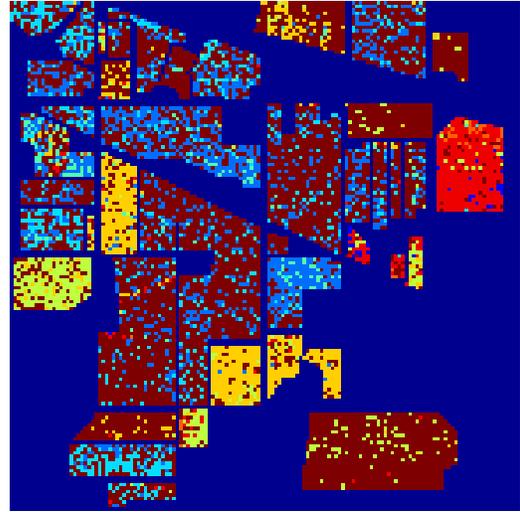
Class map   Ground truth map: 50.9%

Figure 4.14: LE results for Indian Pines
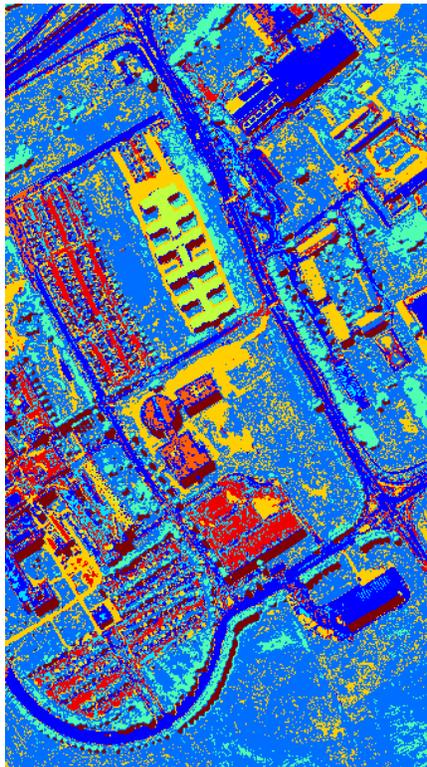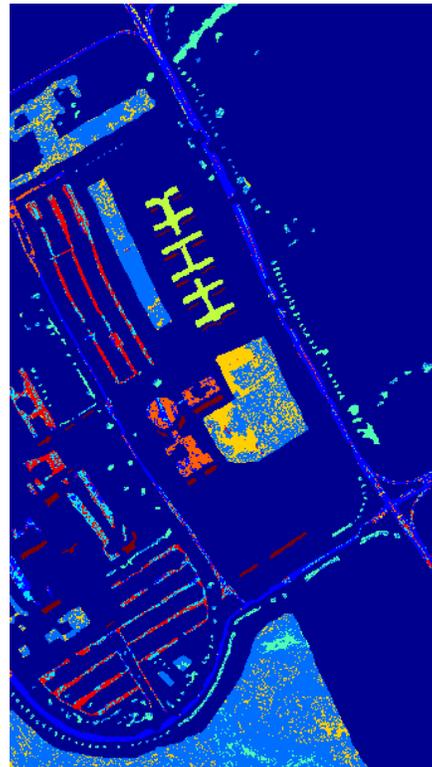
Class map   Ground truth map: 73.6%

Figure 4.15: LE results for Pavia University

Chapter 5

Extensions

5.1   Spatial/Spectral Analysis

Spatial information is inherently provided in the datacube, however, it wasn't until recently that analyst started taking advantage of it. The problem of spatial spectral fusion has emerged as an important and active area of research in HSI analysis. It provides a natural means of smoothing results locally, yielding better classification.

Some recent work includes the use of wavelet packets [3], modified distance metrics for graph construction [44], combining spectral classification with image segmentation results [53], and adding spatial information in spectral band grouping [38].

We show that there is a natural way of introducing spatial information into the LE algorithm that yields results comparable to the state of the art.

In this section, we work only with the Pavia University dataset, whose specifics can be found in Section 4.2.2. Unless otherwise stated, the parameters and methodologies will be the same.

5.1.1   Pure Spatial LE

First, we must note that too much spatial information is a bad thing. Too much spatial information will completely wash out the spectral characteristics of the image. However, adding spatial information will typically raise the classification percentage. Take, for example, the analysis of Pavia University using Laplacian Eigenmaps with only spatial information.

Definition 5.1.1. The *spatial metric* is given by

$$\|x_i - x_j\|_s = \|s(i) - s(j)\|,$$

where the mapping $s : \{1, \dots, MN\} \to \{1, \dots, M\} \times \{1, \dots, N\}$ provides the index for the pixel x_i, and $\| \cdot \|$ is the usual ℓ^2-norm. We will write $\|x_i - x_j\|_f$ for the *spectral metric*, i.e. the usual distance between two points in the datacube, as spectral vectors in $\mathbb{R}^D$.

Pure spatial LE is done simply by replacing the spectral metric with the spatial metric during the constructions of the kNN-graph and the Laplacian. Looking at the ground truth map in Figure 5.1, the overall accuracy of classification has increased significantly. However, that is clearly false progress as the class map shows.
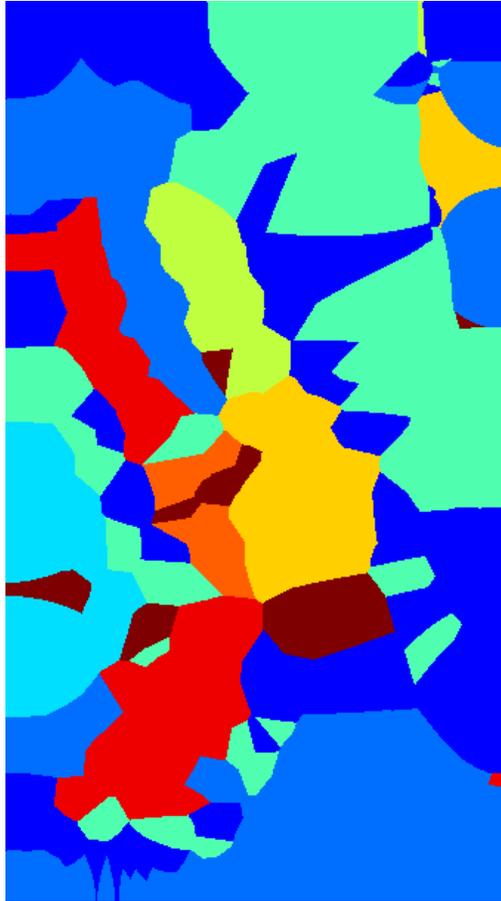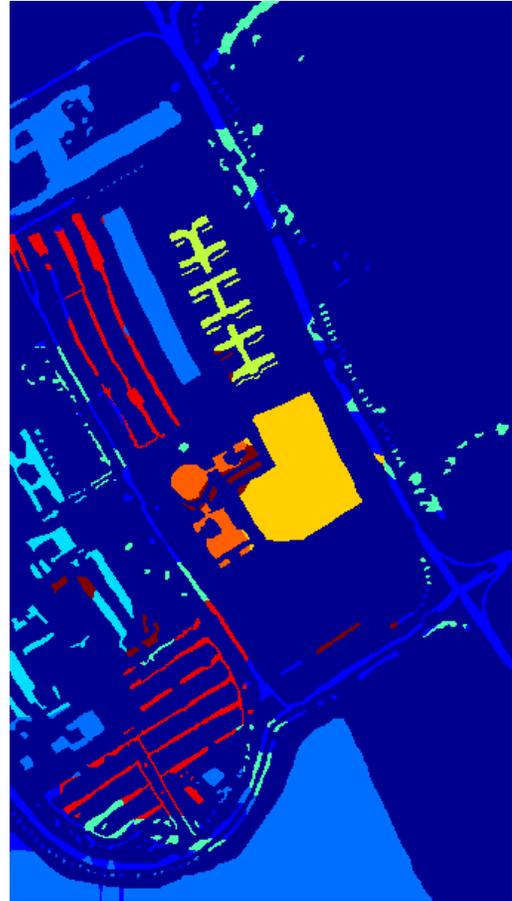
Class map	Ground truth map: 89.3%

Figure 5.1: Purely spatial LE results for Pavia University

The conflicting messages given by the two images in Figure 5.1 is due to the spatial bias present in the ground truth. In the absence of complete ground truth, to truly test spatial/spectral methods, one needs ground truth that is an emulsion of classes. The datasets that are publicly available tend to have ground truth classes separated into distinct spatial clumps.

Thus, we will proceed by adding spatial information a little at a time. First, we will introduce it into the Laplacian, forming *spatial/spectral operators*. Then we will introduce it into the metric, so that the underlying geometry is now a *spatial/spectral manifold*.

5.1.2 Spatial/Spectral Operators

In this section, we introduce spatial information into our analysis by modifying the Laplacian in three different ways. Let G be the usual kNN-graph, constructed using the spectral metric. Let L_f be the usual Laplacian obtained from this graph, using the spectral metric in 4.4. Let L_s be the Laplacian obtained by substituting the spatial metric in the definition of the weights. These Laplacians each depend on their separation parameter. Thus, we define

$$L_1(\sigma, \eta) = L_f(\sigma) \cdot L_s(\eta) \tag{5.1}$$

$$L_2(\sigma, \eta) = L_f(\sigma) + L_s(\eta), \tag{5.2}$$

$$L_3(\sigma, \eta) = G \cdot (L_f(\sigma) \times L_s(\eta)), \tag{5.3}$$

where $\cdot$ is entrywise multiplication and $\times$ is usual matrix multiplication.

We fix $\sigma = 1$, and let $\eta = .2$. The results of the analysis can be found in Figures 5.2, 5.3 and 5.4.
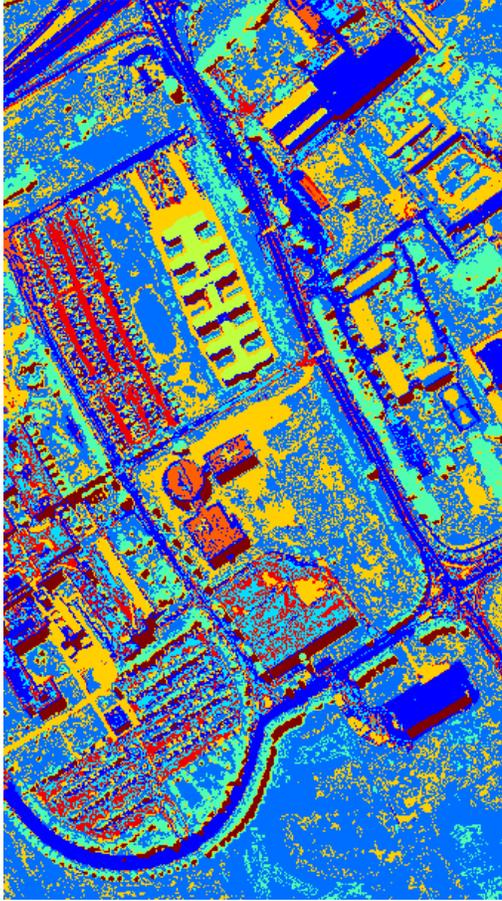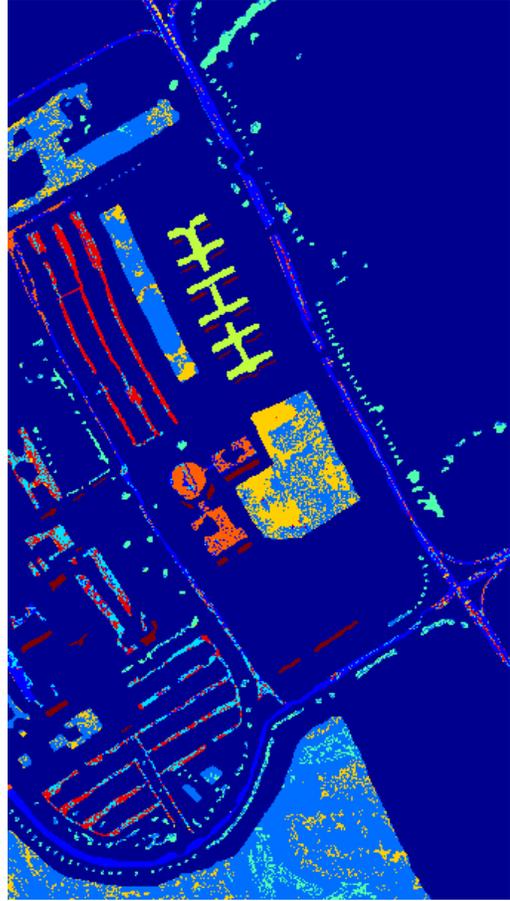
Class map   Ground truth map: 74.7%

Figure 5.2: Classification results for $L_1(1, .2)$
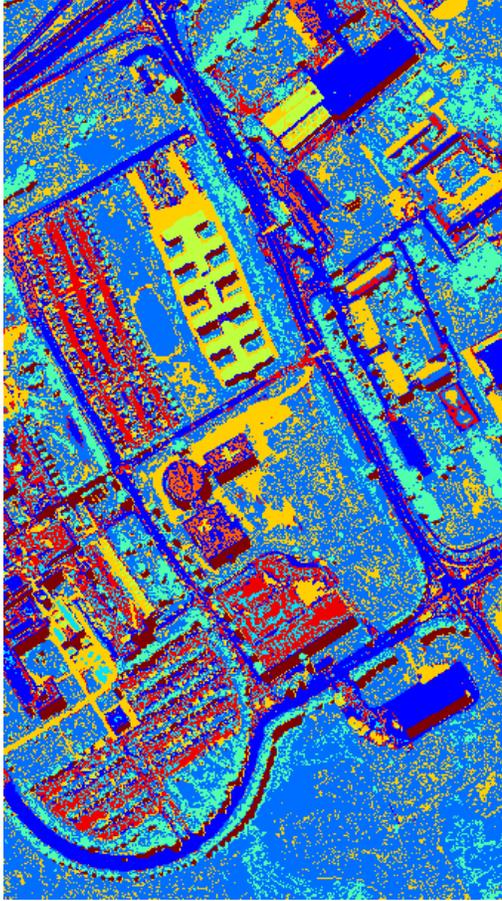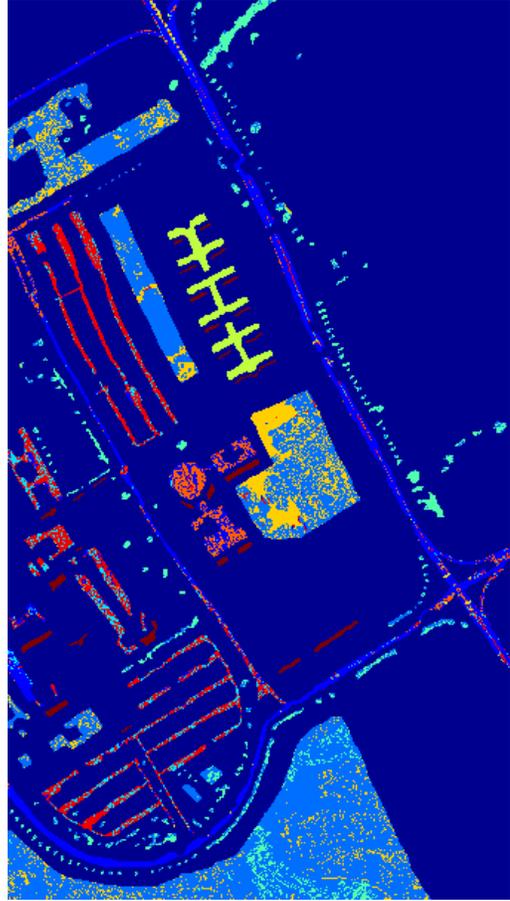
Class map   Ground truth map: 72.6%

Figure 5.3: Classification results for $L_2(1, .2)$
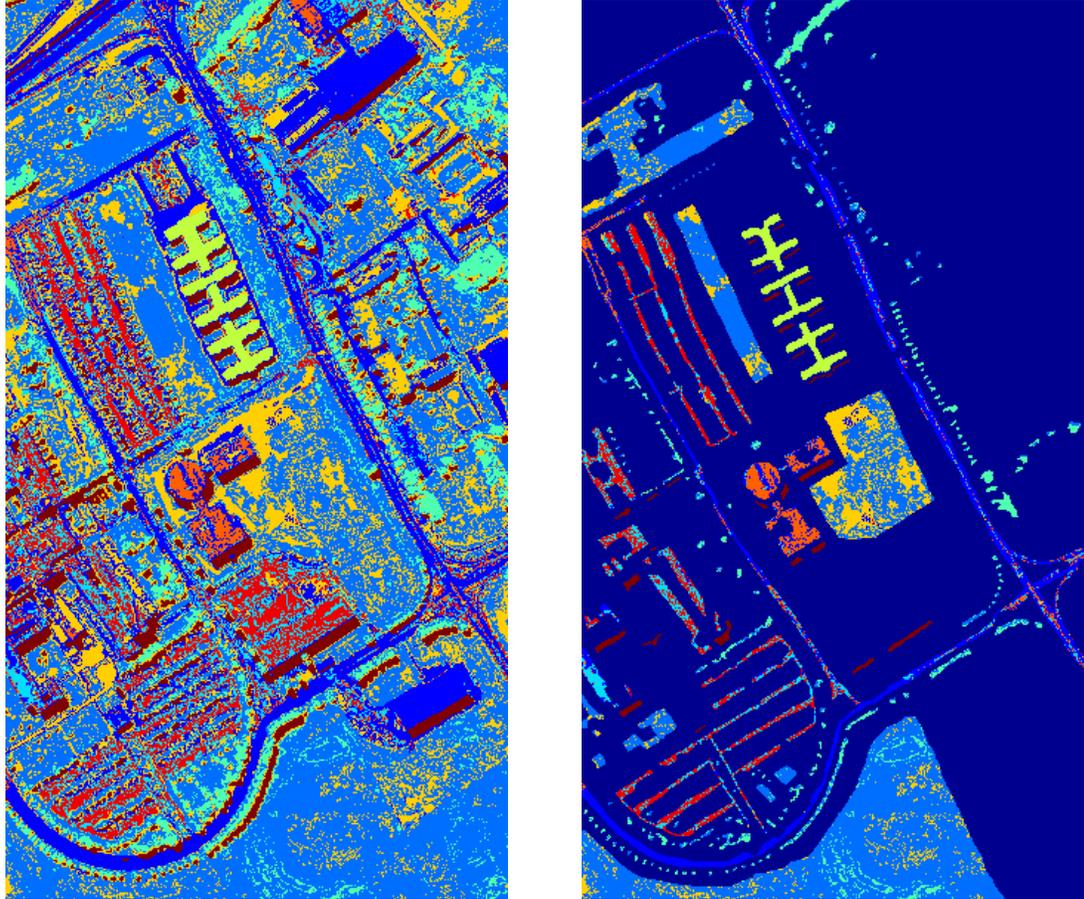
Class map   Ground truth map: 74.1%

Figure 5.4: Classification results for $L_3(1, .2)$

The results suggest that adding spatial information in the operator alone will yield only a slight improvement in classification.

5.1.3 Spatial/Spectral Manifold

Definition 5.1.2. Let $\gamma > 0$. We define the *spatial/spectral metric* on X as follows:

$$\|x_i - x_j\|_\gamma = \left[\|x_i - x_j\|_f^2 + \gamma\|x_i - x_j\|_s^2\right]^{1/2}.$$

Two pixels in our image are close with respect to $\|\cdot\|_\gamma$ if and only if they are

close both spatially and spectrally. The parameter γ allows the user control over how much spatial information to include in the analysis.

Using the spatial/spectral metric in the determination of the kNN-graph and in 4.4 amount to changing the geometry of X, so that it is now a spatial/spectral manifold. This is clearly a more significant adjustment, and so we expect a more significant change in the results.

Again, we let $\sigma = 1$, and computed results for $\gamma \in \{20, 25, 30, 35, 40\}$:

Class map   Ground truth map: 96.9%
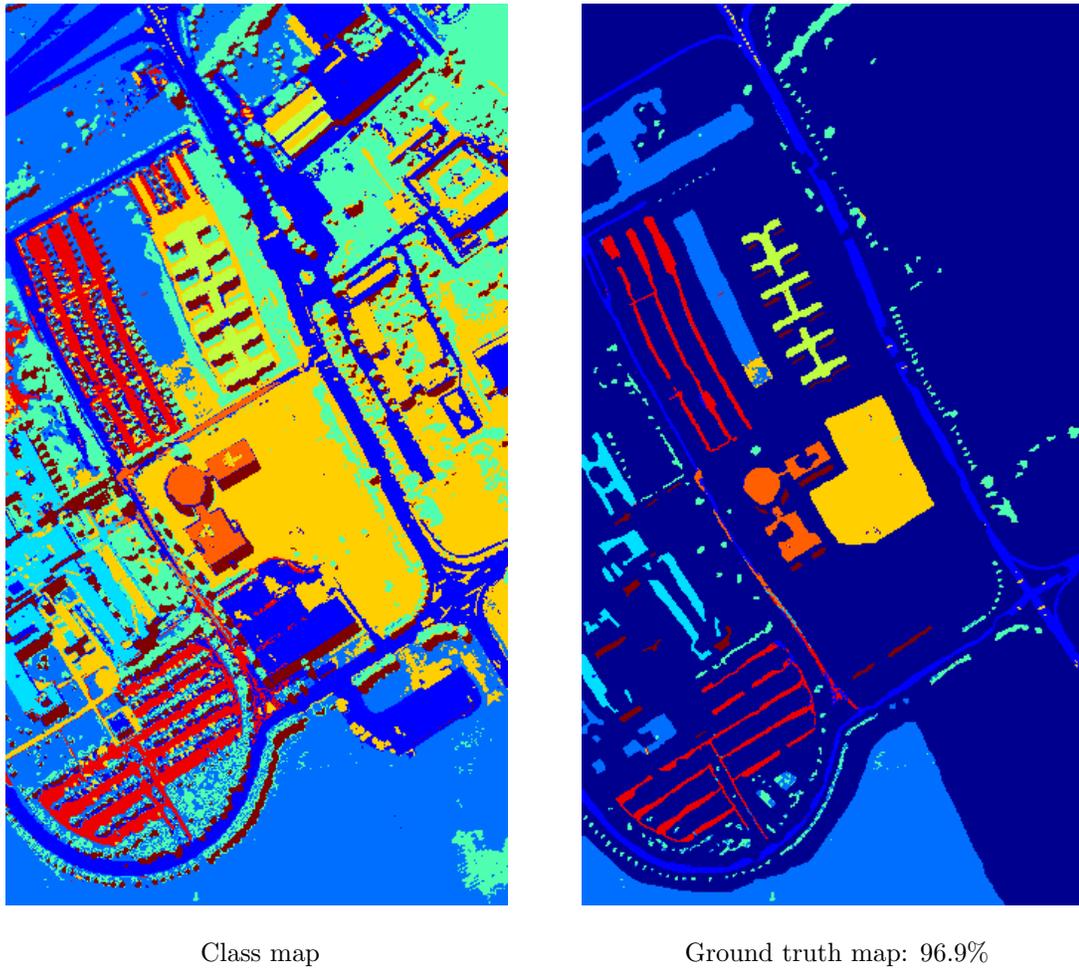
Figure 5.5: Classification results for $\gamma = 20$
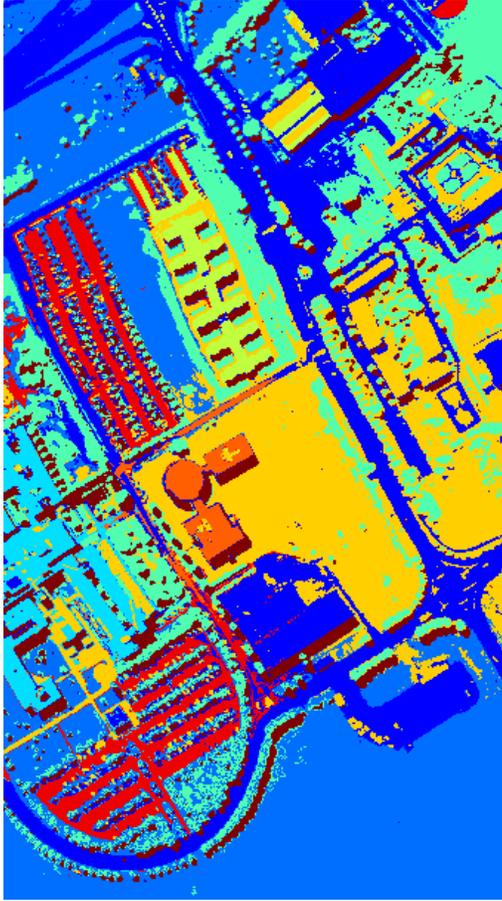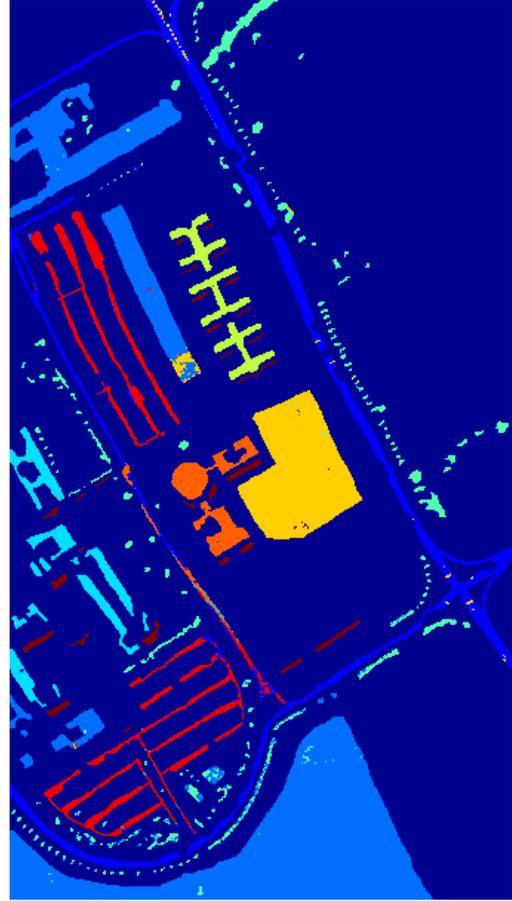
Class map   Ground truth map: 97.4%

Figure 5.6: Classification results for $\gamma = 25$
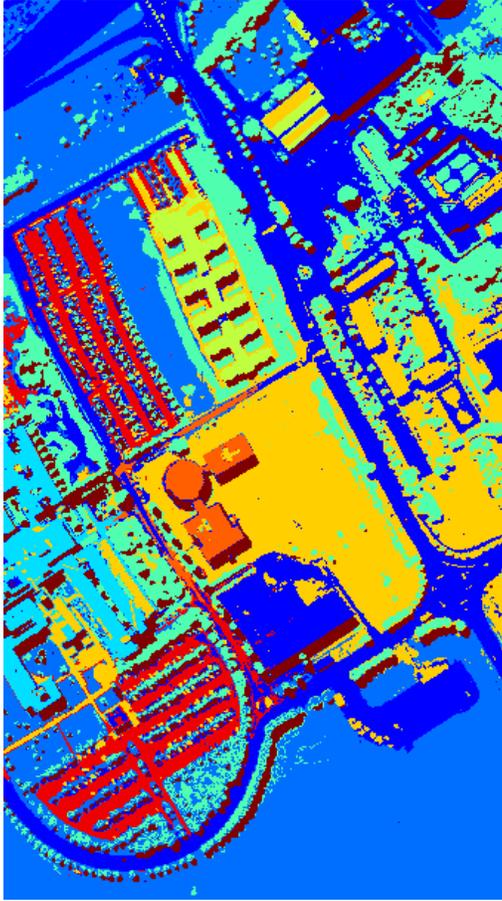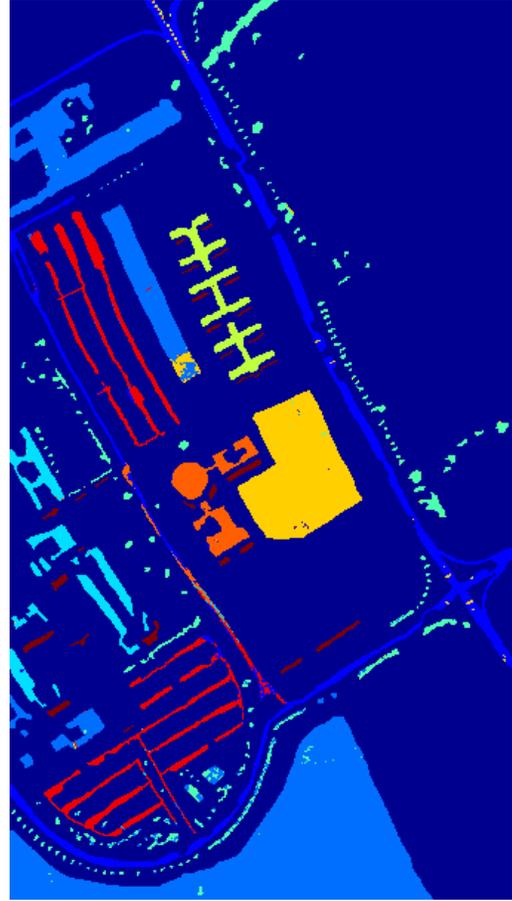
Class map   Ground truth map: 97.6%

Figure 5.7: Classification results for $\gamma = 30$
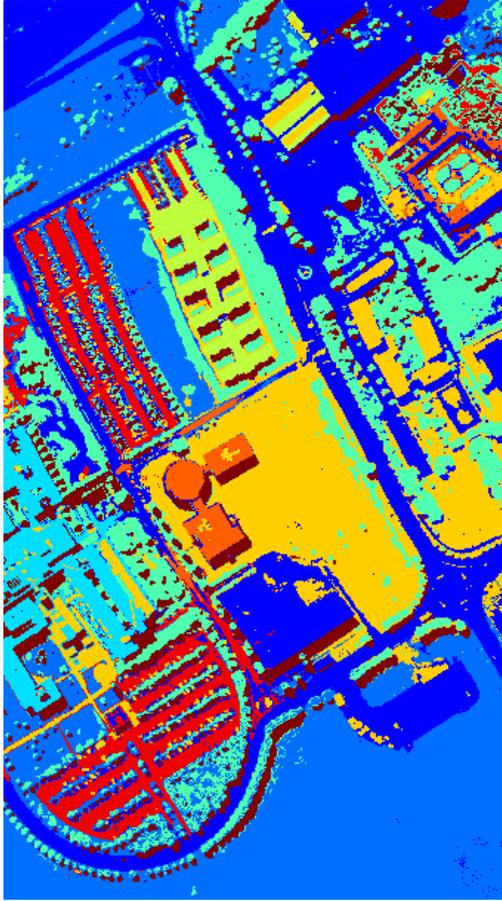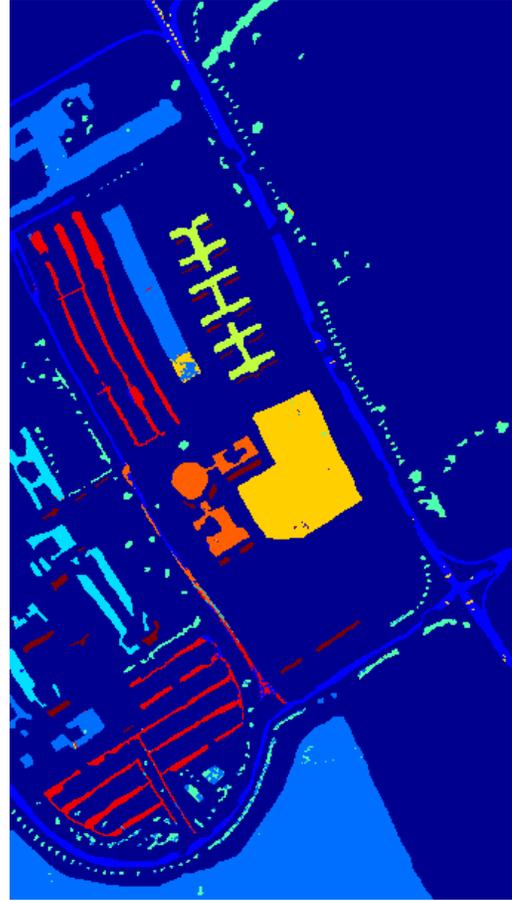
Class map   Ground truth map: 97.9%

Figure 5.8: Classification results for $\gamma = 35$
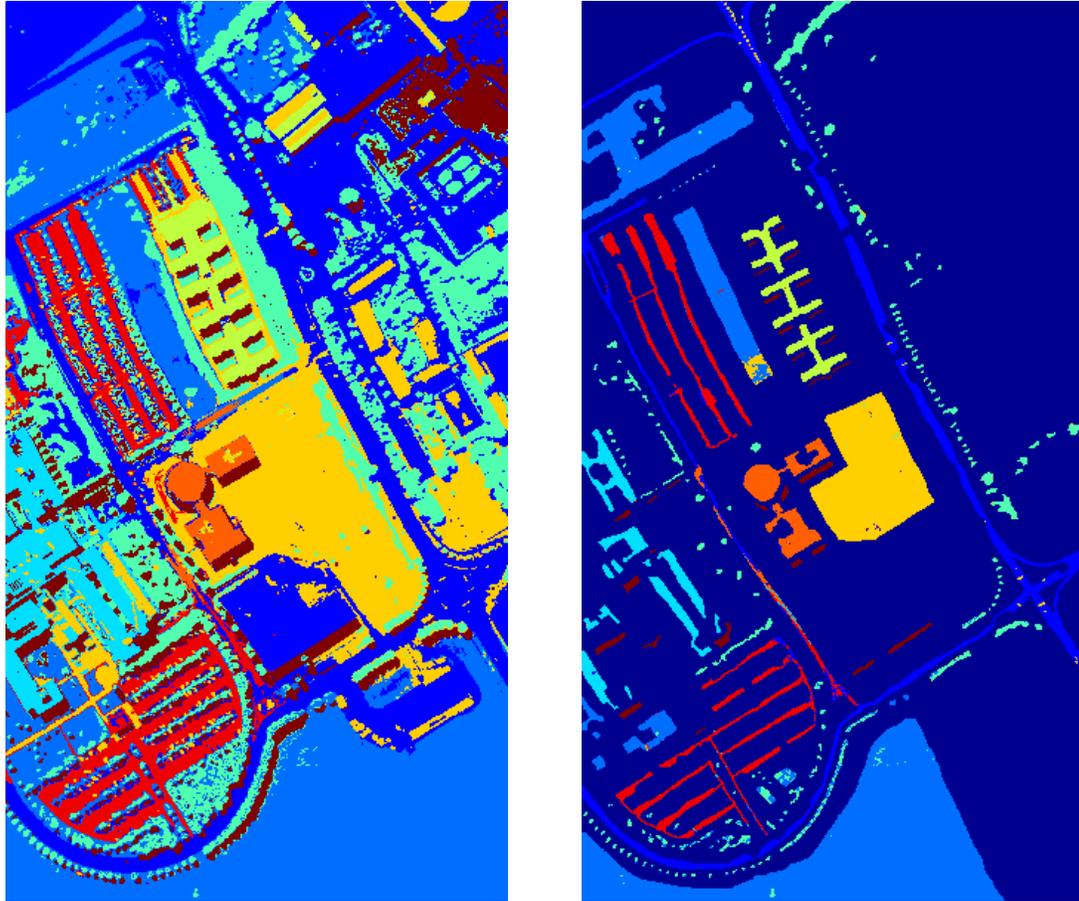
Class map   Ground truth map: 97.9%

Figure 5.9: Classification results for $\gamma = 40$

The results indicate a significant improvement over standard LE. Moreover, the class maps indicate that the progress is legitimate.

5.2   Schrödinger Eigenmaps

In [11], Czaja and Ehler introduced Schrödinger Eigenmaps (SE), a means of generalizing LE by introducing a potential, V, on the graph. We will use SE in our analysis of HSI images. We will provide a proof of concept that certain potentials

provide a semisupervised method for isolating features in an image. We will also show that SE provides an interesting platform for which the spectral clustering of LE can be fused with other clustering techniques. In particular, we will show that this leads to significant improvement in the analysis of Indian Pines.

5.2.1 Introduction

Definition 5.2.1. Let V be a symmetric, positive semi-definite $N \times N$ matrix and let $\alpha > 0$. The operator

$$E_\alpha = L + \alpha V, \qquad (5.4)$$

is called a *graph Schrödinger operator*.

The weight α indicates the relative significance of the potential with respect to the Laplacian operator.

We replace (4.3) with the minimization problem

$$\underset{y^T D y = 1}{\arg \min} = y^T E_\alpha y. \qquad (5.5)$$

The graph Schrödinger operator is analogous to the Schrödinger operator defined on a manifold. In fact, it is more than analogous. In his thesis [23], Halevy extended Belkin and Niyogi's result on pointwise and spectral convergence to the Schrödinger operator.

We consider two types of potentials: *barriers* and *clusters*.

Barriers

Definition 5.2.2. A *barrier potential*, V, is a nonnegative diagonal matrix. The *barrier nodes* are the vertices, i, for which $V(i) > 0$.

Similar to (4.1), for barriers we have

$$y^T E_\alpha y = y^T (L + \alpha V) y$$

$$= \frac{1}{2} \sum_{i,j} (y_i - y_j)^2 W_{ij} + \alpha y^T V y$$

$$= \frac{1}{2} \sum_{i,j} (y_i - y_j)^2 W_{ij} + \alpha \sum_i V(i) y_i^2 \qquad (5.6)$$

Thus (5.5) is equivalent to solving

$$\underset{y^T D y = 1}{\arg\min} \frac{1}{2} \sum_{i,j} (y_i - y_j)^2 W_{i,j} + \alpha \sum_i V(i) y_i^2 \qquad (5.7)$$

The diagonal potential introduces a penalty term at each nonzero entry, which creates a well at the corresponding barrier nodes. More precisely, let y_α be a solution to (5.7) for $\alpha > 0$. If $V(i) \neq 0$, then

$$|y_\alpha(i)|^2 \leq \frac{C}{\alpha V(i)}. \qquad (5.8)$$

Thus, as $\alpha \to \infty$, $y_\alpha(i) \to 0$. The embedding must preserve local geometry, however, so the neighbors of these vertices are also pulled toward 0.

Note that if G is connected, E is nonsingular. Hence there are no zero solutions to worry about in this case. Again, the minimization problem can be realized by solving a generalized eigenproblem:

$$E_\alpha v = \lambda D v. \qquad (5.9)$$

Clusters

Definition 5.2.3. Clusters Let $\iota = \{i_1, \ldots, i_m\}$ be a collection of vertices. A *cluster potential* over ι is defined by taking V to have zeros everywhere except for $V[\iota, \iota]$, the submatrix defined by ι. For this submatrix, define

$$V[\iota, \iota] := \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}. \tag{5.10}$$

If we define a graph, $\widetilde{G}$, on our collection of vertices by drawing an edge between i_k and i_{k+1}, we have

$$\widetilde{G} = \begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ & & & & 1 & 0 \end{pmatrix}. $$

Thus, $V[\iota, \iota] = \widetilde{D} - \widetilde{G}$ is a graph Laplacian. It follows that (5.5) is equivalent to

$$\arg\min_{y^T D y = 1} \frac{1}{2} \sum_{i,j} (y_i - y_j)^2 W_{i,j} + \alpha \sum_{k=1}^{m-1} (y_{i_k} - y_{i_{k+1}})^2. \tag{5.11}$$

The penalty term has the effect of sequentially identifying points in ι; as $\alpha \to \infty$,

$$\sum_{k=1}^{m-1} (y_{i_k} - y_{i_{k+1}})^2 \leq \frac{C}{\alpha} \to 0. \tag{5.12}$$

Again, the embedding must preserve local geometry, so neighbors of ι are pulled inward as well. Thus, the potential forces an increased amount of clustering around ι.

Constant vectors are zeros for E_α. Since both L and V are Laplacians, if G is connected, it follows that they are the only zeros. In fact, since V is assumed to positive semidefinite, $\ker(y^T E_\alpha y) \subset \ker(y^T L y)$, for any V. Thus, we may ignore the zero eigenvalue and look for the d smallest nonzero eigenvalues.

5.2.2 Random Barrier Potentials

As in Section 4.2.2, we fix $k = 20$ for the number of nearest neighbors, $\sigma = 1$, and $d = 50$ for the number dimensions in the embedding.

A naive barrier potential on each class of the ground truth can be defined as follows.

Definition 5.2.4. Choose $0 < p < 1$. For each class k (there are 16 for Indian Pines), let N_k be the number of ground truth for that class. Define a barrier potential V_k by randomly selecting pN_k of the ground truth from the k-th class. Define V^k to be 1 at each of the selected nodes. Given $\alpha > 0$, define

$$E_\alpha^k = L = \alpha V_k.$$

We chose $p = .2$ and analyzed the embeddings generated by E_α for $\alpha = 100$, $\alpha = 1000$. Vector angle classification was used to classify the results. Because we are collapsing the barrier nodes to a single point, we omit them from the classification; 10% from each class of the remaining ground truth is selected for training. Recall,

our baseline comparison is the overall accuracy of using LE on Indian Pines, which was 50.9%.

Overall accuracy of the classification increases by about 10% for each V^k.
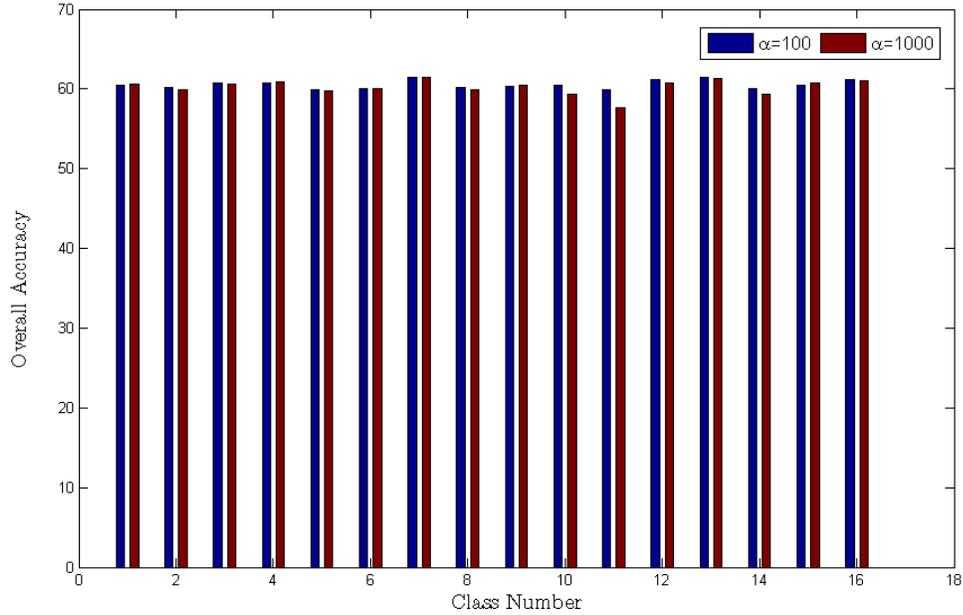
Figure 5.10: Overall accuracy for random barrier potentials

A better demonstration of the possible benefits of barrier potentials, however, is given by studying the embeddings themselves. Figures 5.11 and 5.12 show two and three dimensional projections of the embedding for the class which hosts the barrier potential along with several other classes.
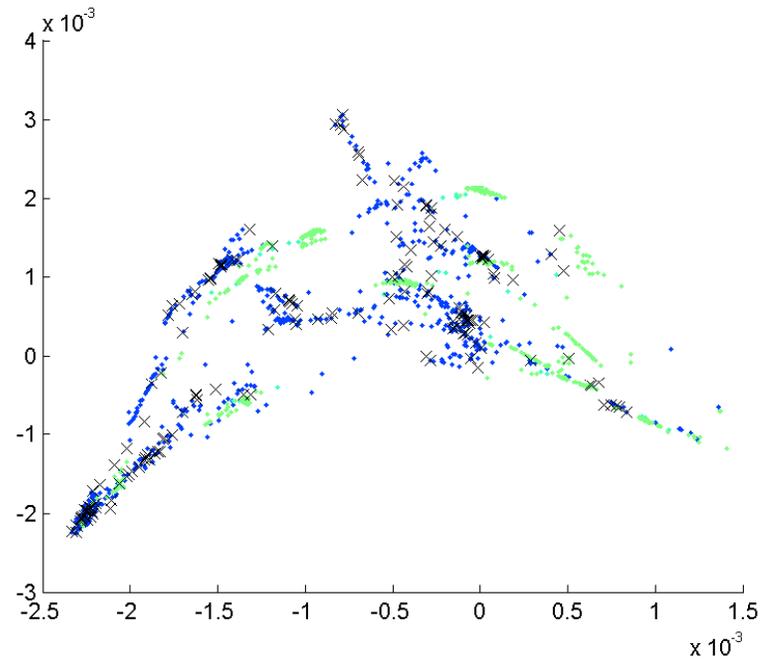
LE
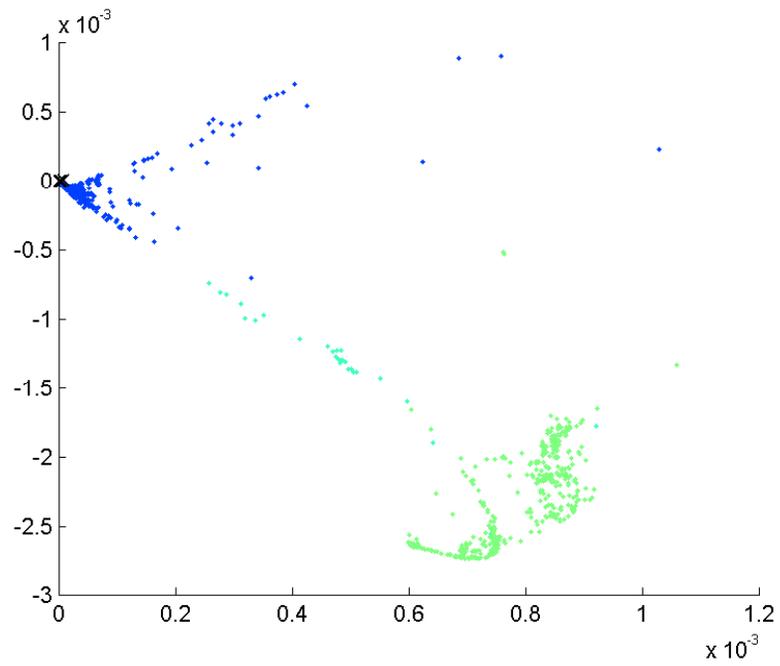
E^3_{1000}

Figure 5.11: Effect of E^3_{1000} in dimensions 1 and 3 for classes 3, 7, and 8
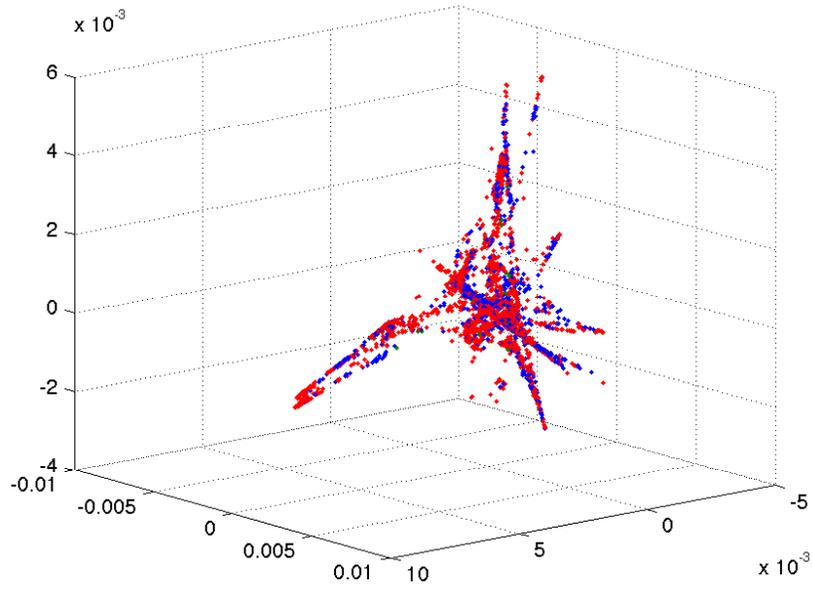
LE
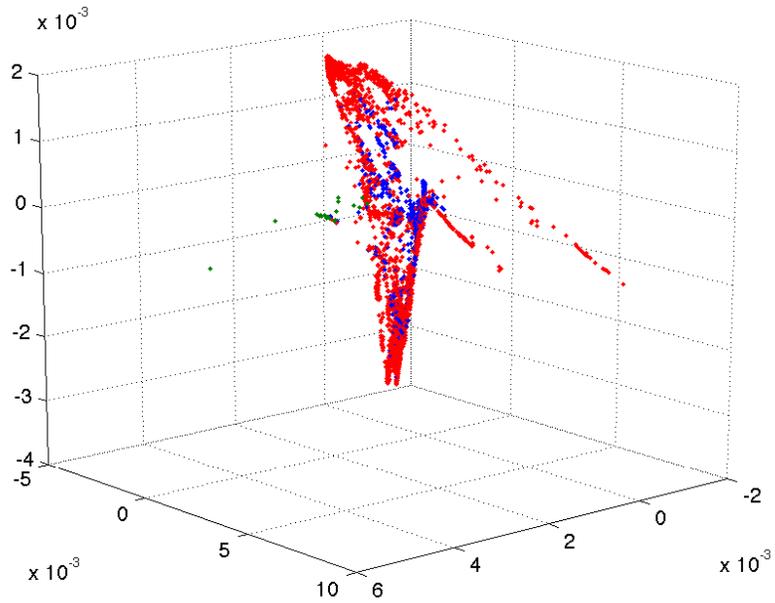
E_{1000}^{7}

Figure 5.12: Effect of E_{1000}^{7} in dimensions 6 , 9, and 13 for classes 7, 10, and 11

Both figures indicate that SE is useful in untangling a particular class from the rest of the data. In Figure 5.11, E^3_{1000} not only separates class 3 from the other two classes, but it also pulls class 7 away from class 8. This suggests that classes 7 and 3 must be spectrally close.

In Figure 5.12, class 7 receives the barrier potential. This class is very small, consisting of only 28 pixels. Thus V^7 is nonzero at only 5 barrier nodes. These points are not pictured in the image.

This example shows that a potential placed at even a few number of points can have a significant effect on the embedding. Although class 7 is buried in classes 10 and 11 in the LE embedding, Figure 5.12 indicates that it is extracted in the SE embedding.

This suggests that barrier potentials could be a very useful semisupervised learning tool in the hands of an expert analyst. Selecting even a few pixels that are of interest and placing a barrier potential at these nodes, it is possible to extract a family of pixels exhibiting a similar behavior.

5.2.3 Cluster Potentials and k-Means Clustering

We define cluster potentials in the following way:

1. Use k-means clustering, with the usual spectral metric, to cluster the hyperspectral image. The clustering is initialized randomly so that the construction of the potential would be completely unsupervised.

2. Using (5.10), define a cluster potential, V_k over each of the $k = 1, \ldots, K$

clusters.

3. Choose an ordering of the points in each cluster.

4. Aggregate the information into one potential by summing over the individual cluster potentials:

$$V = \sum_{k=1}^{K} V_k.$$

In our first experiment, we used two different ordering schemes: spatial and spectral. Spatial ordering is determined by finding the spatial center of the cluster. Points were indexed in order of their proximity to the center. Spectral ordering is done exactly the same, except using the spectral center of the cluster.

We varied the number of clusters, from 10 to 100 by tens and let $\alpha = 100, 1000$. The results indicate significant improvement for all α and K.
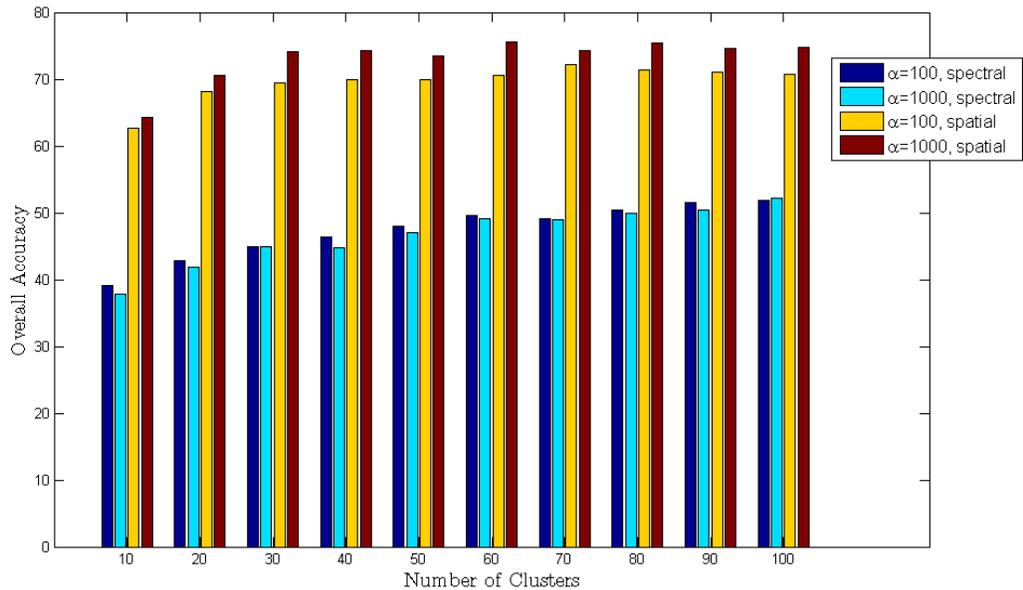
Figure 5.13: Overall accuracy of classification for spatially and spectrally ordered cluster potentials

Spatial ordering gave a significant increase in accuracy over LE, while spectral ordering often performed worse than LE. This indicates that the ordering is an important step in the construction of the cluster potentials.

Motivated by the performance of spatial ordering, we made the following amendment to the construction of the cluster potential. Instead of clustering the entire image, grid the image and separate it into local pieces. Perform kmeans clustering on each piece, order the points inside each cluster, and construct the potential.

We divided India Pines into grids of size $M \times M$ for various M. We also varied K, the number of clusters found in each grid cell. The results for this analysis are given in the Figure 5.14 and 5.15.
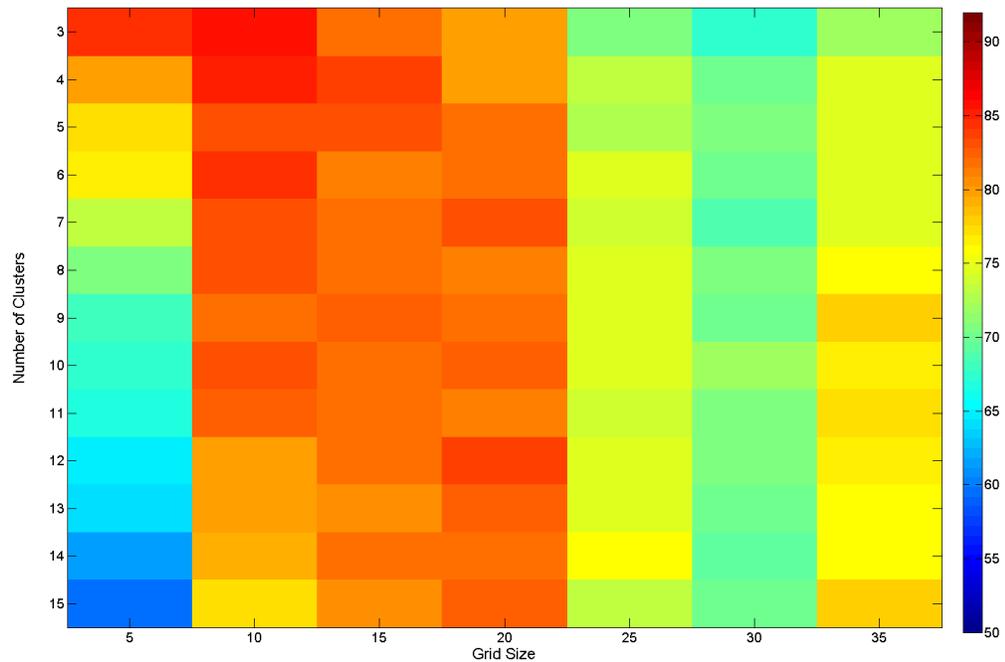
Figure 5.14: Overall accuracy for gridded cluster potentials, $\alpha = 100$
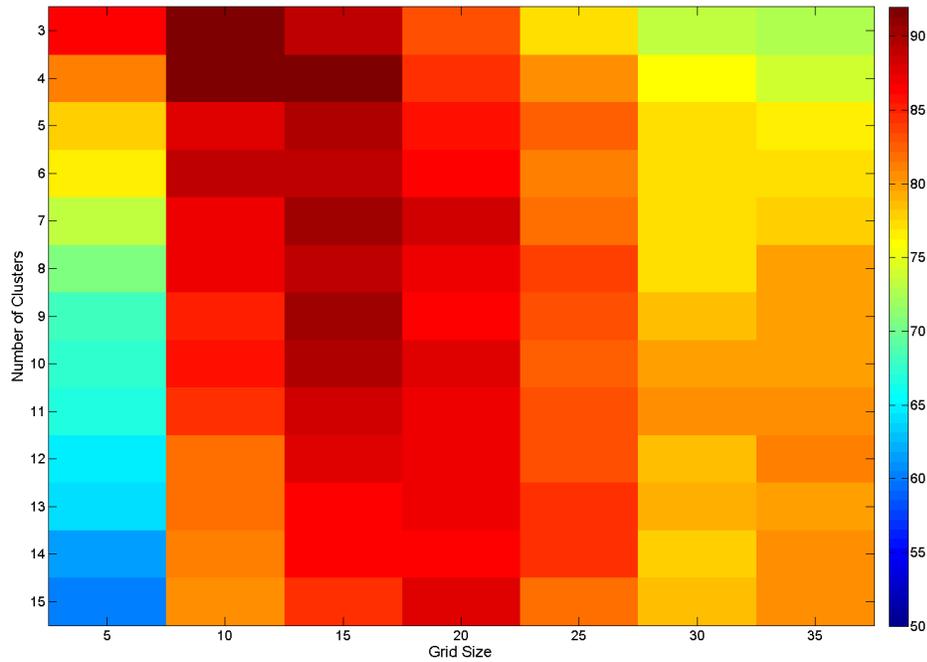
Figure 5.15: Overall accuracy for gridded cluster potentials, $\alpha = 1000$

With the appropriate choice of grid size and number of clusters, the gridded kmeans cluster potential provides a significant improvement over LE. The class maps for the three techniques indicate the same trend. Figures 5.16, 5.17, and 5.18 provide the class map for the best results from each method, with $\alpha = 1000$.

The two dimensional projection of the embedding in Figure 5.19 provides some evidence for the improvement in classification. The cluster potentials control the internal geometry of the individual clusters. As $\alpha \to \infty$, the potential forces the individual clusters to contract. For a large enough number of clusters, this seems to be balanced by the Laplacian, which controls the geometric relationship between the clusters.
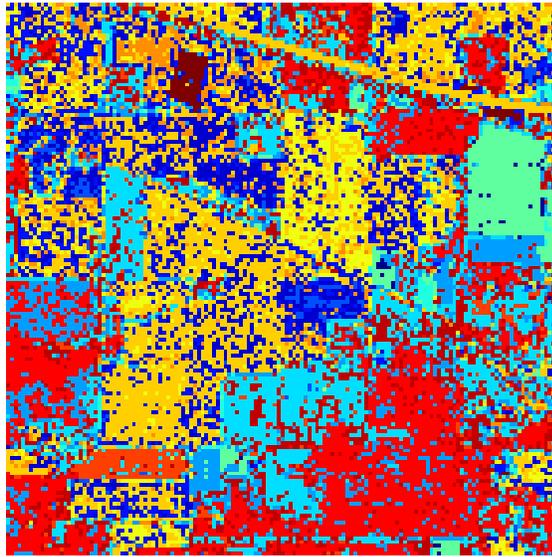
Class map
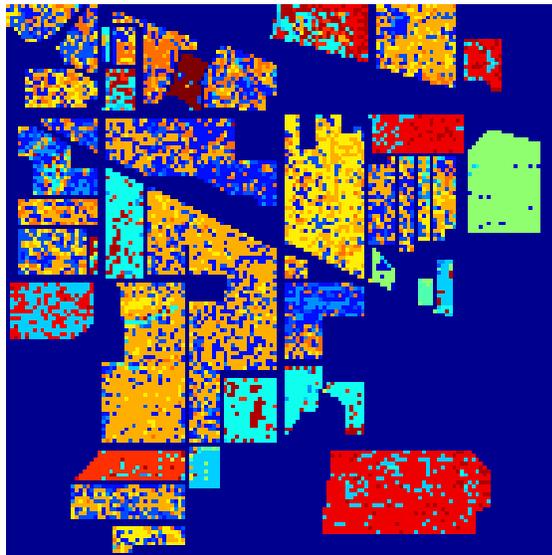
Ground truth map: 52.2%

Figure 5.16: Class map for spectrally ordered cluster potential, 100 clusters

137
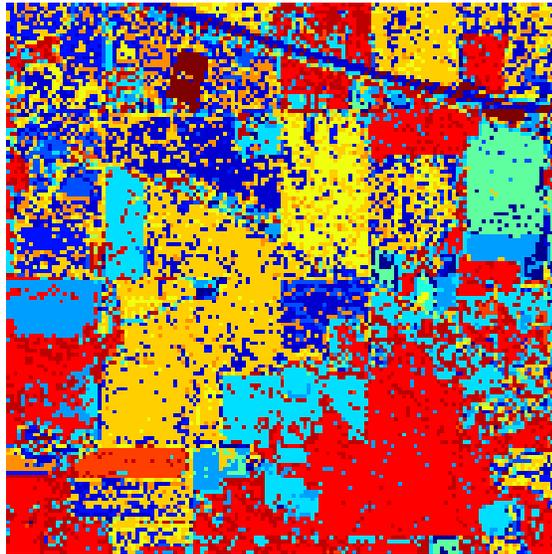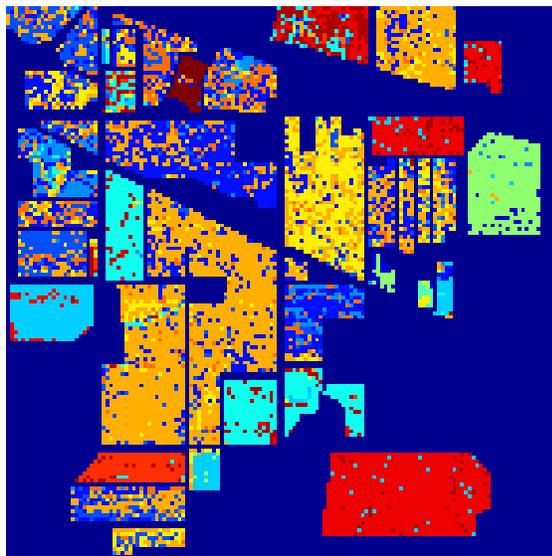
Class map

Ground truth map: 75.6%

Figure 5.17: Class map for spatially ordered cluster potential, 60 clusters
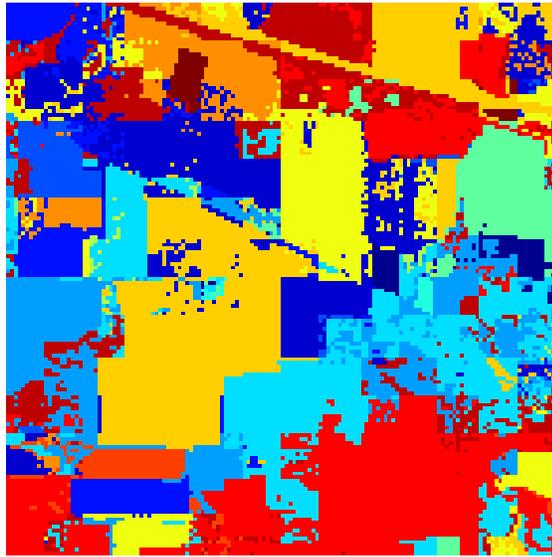
Class map
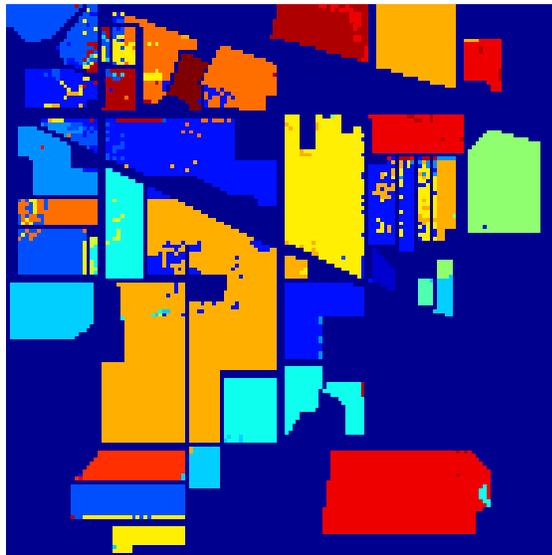
Ground truth map: 91.6%

Figure 5.18: Class map for gridded cluster potential, 15×15 grid, 4 clusters
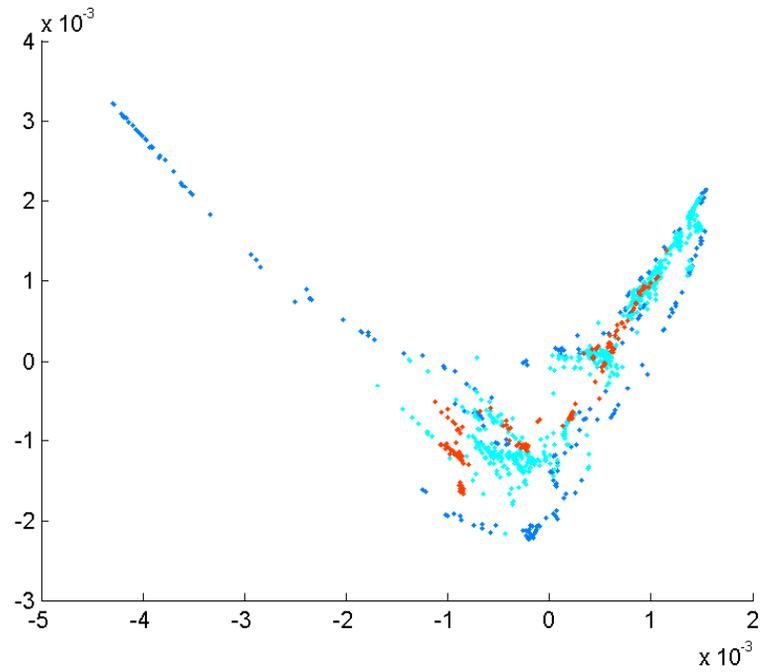
LE
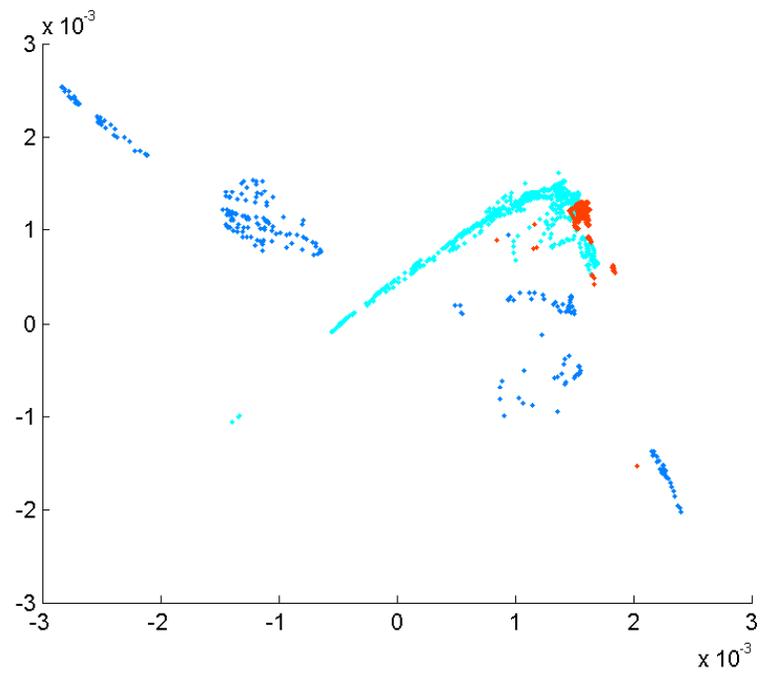
SE

Figure 5.19: Effects of a 4 cluster, 15×15 gridded kmeans potential in dimensions 2

and 3 for classes 4, 6, and 13

Bibliography

[1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6), 2003.

[2] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8), 2008.

[3] J. Benedetto, W. Czaja, M. Ehler, C. Flake, and M. Hirn. Wavelet packets for multi- and hyper-spectral imagery. *Proc. SPIE*, 7535(1), 2010.

[4] John Benedetto. *Harmonic analysis and applications*. CRC, 1997.

[5] John J. Benedetto and Paulo J. S. G. Ferreira, editors. *Modern sampling theory*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2001. Mathematics and applications.

[6] Arne Beurling. *The collected works of Arne Beurling. Vol. 2*. Birkhäuser Boston Inc., Boston, MA, 1989.

[7] L. Bezuglaya and V. Katsnelson. The sampling theorem for functions with limited multi-band spectrum, i. *Z. Anal. Anwendungen*, 12(3), 1993.

[8] Jie Chen, Haw-ren Fang, and Yousef Saad. Fast approximate knn graph construction for high dimensional data via recursive lanczos bisection. *Journal of Machine Learning Research*, 10(2009), 2009.

[9] O. Christensen. *An introduction to frames and Riesz bases*. Birkhauser, 2003.

[10] W. Czaja, G. Kutyniok, and D. Speegle. Beurling dimension of gabor pseudoframes for affine subspaces. *Journal of Fourier Analysis and Applications*, 14(4), 2008.

[11] Wojciech Czaja and Martin Ehler. Schroedinger eigenmaps for the analysis of bio-medical data. *CoRR*, abs/1102.4086, 2011.

[12] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72, 1952.

[13] D.E. Dutkay, D. Han, and E. Weber. Iterative approximations of exponential bases on fractal measures. *Arxiv preprint arXiv:1109.4669*, 2011.

[14] Dorin Ervin Dutkay, Deguang Han, Qiyu Sun, and Eric Weber. On the Beurling dimension of exponential frames. *Adv. Math.*, 226(1), 2011.

[15] Dorin Ervin Dutkay, Deguang Han, and Eric Weber. Bessel sequences of exponentials on fractal measures. *J. Funct. Anal.*, 261(9), 2011.

[16] K.J. Falconer. *Fractal geometry: mathematical foundations and applications.* John Wiley & Sons Inc, 2003.

[17] Bálint Farkas, Máté Matolcsi, and Péter Móra. On Fuglede's conjecture and the existence of universal spectra. *J. Fourier Anal. Appl.*, 12(5), 2006.

[18] Bálint Farkas and Szilárd Gy. Révész. Tiles with no spectra in dimension 4. *Math. Scand.*, 98(1):44–52, 2006.

[19] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications.* Wiley-Interscience, New York, 1999.

[20] Bent Fuglede. Commuting self-adjoint partial differential operators and a group theoretic problem. *J. Functional Analysis*, 16, 1974.

[21] K. Gröchenig and H. Razafinjatovo. On Landau's necessary density conditions for sampling and interpolation of band-limited functions. *Journal of the London Mathematical Society*, 54(3).

[22] Karlheinz Gröchenig, Gitta Kutyniok, and Kristian Seip. Landau's necessary density conditions for LCA groups. *J. Funct. Anal.*, 255(7), 2008.

[23] Avner Halevy. *Extensions of Laplacian Eigenmaps for Manifold Learning.* PhD thesis, University of Maryland, College Park.

[24] X.G. He, C.K. Lai, and K.S. Lau. Exponential spectra in $l^2(\mu)$. *Arxiv preprint arXiv:1110.1426*, 2011.

[25] Edwin Hewitt and Kenneth A. Ross. *Abstract harmonic analysis. Vol. I*, volume 115. Springer-Verlag, Berlin, second edition, 1979.

[26] John E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30, 1981.

[27] Alex Iosevich, Nets Katz, and Terence Tao. The Fuglede spectral conjecture holds for convex planar domains. *Math. Res. Lett.*, 10(5-6), 2003.

[28] Alex Iosevich, Nets Hawk Katz, and Terry Tao. Convex bodies with a point of curvature do not have Fourier bases. *Amer. J. Math.*, 123(1), 2001.

[29] S. Jaffard. A density criterion for frames of complex exponentials. *Michigan Math. J.*, 38(3), 1991.

[30] Palle E. T. Jorgensen and Steen Pedersen. Dense analytic subspaces in fractal L^2-spaces. *J. Anal. Math.*, 75, 1998.

[31] J.-P. Kahane and R. Salem. *Ensembles Parfaits et Séries Trigonométriques.* Hermann, Paris, 1963.

[32] Igor Kluvánek. Sampling theorem in abstract harmonic analysis. *Mat.-Fyz. Časopis Sloven. Akad. Vied*, 15, 1965.

[33] Mihail N. Kolountzakis and Máté Matolcsi. Complex Hadamard matrices and the spectral set conjecture. *Collect. Math.*, (Vol. Extra), 2006.

[34] A. S. Kravchenko. Completeness of the space of separable measures in the Kantorovich-Rubinshteĭn metric. *Sibirsk. Mat. Zh.*, 47(1), 2006.

[35] I. Łaba. Fuglede's conjecture for a union of two intervals. *Proc. Amer. Math. Soc.*, 129(10), 2001.

[36] C.K. Lai. On Fourier frames of absolutely continuous measures. *Journal of Functional Analysis*, 2011.

[37] H. J. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.*, 117, 1967.

[38] M.A. Lee, L.M. Bruce, and S. Prasad. Concurrent spatial-spectral band grouping: Providing a spatial context for spectral dimensionality reduction. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, june 2011.

[39] N. Lev. Riesz bases of exponentials on multiband spectra. *Arxiv preprint arXiv:1101.3894*, 2011.

[40] Y.I. Lyubarskii and A. Rashkovskii. Complete interpolating sequences for fourier transforms supported by convex symmetric polygons. *Arkiv för Matematik*, 38(1).

[41] Yurii I. Lyubarskii and Kristian Seip. Sampling and interpolating sequences for multiband-limited functions and exponential bases on disconnected sets. *J. Fourier Anal. Appl.*, 3(5), 1997. Dedicated to the memory of Richard J. Duffin.

[42] B. Matei and Y. Meyer. Simple quasicrystals are sets of stable sampling. *Complex Variables and Elliptic Equations*, 55(8-10), 2010.

[43] M. Matolcsi. Fuglede's conjecture fails in dimension 4. *Proceedings of the American Mathematical Society*, 133(10).

[44] A. Mohan, G. Sapiro, and E. Bosch. Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 4, 2007.

[45] Joaquim Ortega-Cerdà and Kristian Seip. Fourier frames. *Ann. of Math. (2)*, 155(3), 2002.

[46] Raymond E. A. C. Paley and Norbert Wiener. *Fourier transforms in the complex domain*, volume 19 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1987.

[47] D. Priest and W.M. Arkin. A hidden world growing beyond control, July 19 2010. The Washington Post.

[48] Jayakumar Ramanathan and Tim Steger. Incompleteness of sparse coherent states. *Appl. Comput. Harmon. Anal.*, 2(2), 1995.

[49] Hans Reiter and Jan D. Stegeman. *Classical harmonic analysis and locally compact groups*, volume 22 of *London Mathematical Society Monographs. New Series*. The Clarendon Press Oxford University Press, New York, second edition, 2000.

[50] Walter Rudin. *Fourier analysis on groups*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1990.

[51] Kristian Seip. On the connection between exponential bases and certain related sequences in $L^2(-\pi, \pi)$. *J. Funct. Anal.*, 130(1), 1995.

[52] Terence Tao. Fuglede's conjecture is false in 5 and higher dimensions. *Math. Res. Lett.*, 11(2-3):251–258, 2004.

[53] Yuliya Tarabalka, Jón Atli Benediktsson, and Jocelyn Chanussot. Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47, August 2009.

[54] Robert M. Young. *An introduction to nonharmonic Fourier series*, volume 93 of *Pure and Applied Mathematics*. Academic Press Inc., New York, 1980.