Two Multiresolution Frameworks on Graphs

Shashank Sule

April 15, 2020

Submitted to the
Department of Mathematics and Statistics
of Amherst College
in partial fulfillment of the requirements
for the degree of
Bachelor of Arts with honors

Faculty Advisor: Professor Karamatou Yacoubou Djima

Contents

List of Figures

List of Notation

$\mathbb{R}$: The set of Real numbers

$\mathbb{C}$: The set of Complex numbers

$\mathbb{Z}$: The set of Integers

$\mathbb{Z}_N$: The set of Integers Modulo $N \in \mathbb{Z}$

$\mathbb{N}_0$: The set of Whole numbers

$[n] = \{1, \ldots, n\}$

$\binom{[n]}{k}$: The set of all k-element subsets of $[n]$

$M_n(\mathbb{R})$: The set of all $n \times n$ real matrices

$\mathbb{R}^n = \{(x_1, \ldots, x_n)^\top \mid x_i \in \mathbb{R}\}$

$\mathbf{1}_n = x \in \mathbb{R}^n$ such that $x_i = 1$ for every $i : 1 \leqslant i \leqslant n$

χ_Ω: The characteristic function on $\Omega \subset \mathbb{R}^n$

List of Definitions

Acknowledgements

My mother and I often muse that as a child I was never "destined" for mathematics. What changed? I began college as a Chemistry and Economics major but in my second year I took a class with Prof. Yacoubou Djima on Fourier Analysis. I think it was the right combination of teacher and subject that made me fully re-evaluate my preferences and decide to pursue mathematics. It was a subject too beautiful not to pursue and in Prof. Yacoubou Djima I had found a mentor who motivated me on an academic and personal level. So my first point of gratitude is towards her for being such an awesome thesis advisor. I would have never discovered this subject but for her and her guidance has contributed heavily towards this thesis. She somehow knew exactly when to encourage or caution me and has been a large part of my excitement about this project. She has also been incredibly forgiving about my tardiness with deadlines, my late requests for recommendations, and the down-to-the wire submissions that I am often prone to.

I also want to thank the present and past faculty of the Mathematics and Statistics department at Amherst who have made learning mathematics such a joy. In particular, I would like to mention Prof. Benedetto for her infectious enthusiasm, Prof. Contreras, Alvarado, and Amalia for their hilarious conversations and graduate school advice, Prof. Daniels for his eternal wisdom and Prof. Sosa for showing me that perseverance and close attention is the way to succeed at mathematics. I loved all my interactions with the Mathematics department. They are some of my fondest memories at Amherst.

I thank my parents for being my moral compasses and support especially as I went through some plucky times during the course of this thesis. Also, to my brother and my sister-in-law for their wisdom and for being such fun companions. Finally, I thank the set of all my friends who have kept me grounded in curiosity, humour, and learning.

Abstract

In this thesis we study the mathematical properties of the Graph Fourier basis (GFB) and the applications of diffusion maps from the viewpoint of multi-scale analysis. In Chapter 1 we introduce multiresolution analyses (MRA) and the mathematical foundations of image processing, and motivate the paradigm of graphs through an example from feature detection. In Chapter 2, we provide motivation for deriving the GFB from the graph Laplacian by demonstrating that eigenfunctions of the Euclidean Laplacian on $[-\pi, \pi]$ exhibit multi-scale properties. In the latter half of the chapter we examine these properties in the GFB context. In Chapter 3 we introduce diffusion maps from [2] and discuss their applications from the perspective of low-dimensional embeddings and multi-scale properties. To end the thesis, we propose Neumann maps as novel constructions that improve on diffusion maps in problems which involve embedding submanifolds from ambient low-dimensional manifolds.

Chapter 1

Introduction

With applications as broad and critical as medical imaging or troubling deep fakes, image processing is one of the most rapidly evolving fields of science, engineering and mathematics. In mathematics, Harmonic Analysis has been at the forefront of developments in this field, treating images as functions to be decomposed into pieces useful for various problems. More precisely, an image can be represented by a set of finite samples $\{f(x_i, y_j)\}_{1 \leqslant i,j \leqslant n}$ of a function $f : [0,1]^2 \mapsto \mathbb{R}$, where (x_i, y_j) is a *pixel* location and $f(x_i, y_j)$ is its *luminescence*. This "digitized" approach towards images allows resolution, or detail, to be encoded by *sampling*: one can sharpen or blur an image by altering the discretization of the unit square, or in other words, representing the image with more or less pixels. In 1989, Mallat [21] challenged the practice of tying resolution to sampling. He argued that sampling is inherently *non-adaptive*: some regions within an image can require denser sampling than others to obtain a given resolution. In the same paper he proposed *Multiresolution Analyses* (MRAs) as an alternative framework that encoded resolution by the projections of *finite-energy* or L^2-functions f onto a sequence of subspaces $\{V_j\}$ converging monotonically to $L^2([0,1]^2)$, which we define formally in the next section. Mallat's work was a landmark result in the rapidly growing field of Wavelet Theory in the 20th century. Decades earlier, Haar [16] had found a basis for $L^2(\mathbb{R})$ using functions of compact support which suggested the possibility of sparse representations in the time domain. Grossmann and Morlet took this project of finding orthonormal bases for $L^2(\mathbb{R})$ further: they proposed *wavelets* $\psi(x)$ whose translations and dilations $\{s^{1/2}\psi(s(x-t))\}_{s,t \in \mathbb{R}^+}$ spanned $L^2(\mathbb{R})$ orthogonally [15]. Yves Meyer proposed an example where the parameters s, t were *integers*: this came to be

known as the *Meyer wavelets* [22]. MRAs show that the dilation and translation properties of wavelets describe resolution naturally; they have since set the standard in image and signal processing. In her work *Ten lectures on wavelets* [9], Daubechies generalized Haar's example to propose compactly supported wavelets of arbitrary smoothness (unlike Haar's which is not even continuous). The JPEG 2000 algorithm uses Daubechies' wavelets to compress and store images in computers.

Currently, however, there is another paradigm of signal processing in fashion: the paradigm of graphs. Due to the emergence of Spectral Graph Theory [6], signal processing has found applications to problems on random sensor networks, high-dimensional data, and computer vision [27] where the digitization perspective can often be limited; it is instead more efficient to model pixels as vertices of a graph and a signal as a function on those vertices. The graph theoretic approach turns out to be quite useful in solving a wide-range of problems from cluster detection and non-linear dimensionality reduction to more classical applications like denoising and feature detection [23]. Consequently, there has been recent interest in defining Multiresolution analyses on graphs [18, 19, 17, 7, 8, 1]. This thesis is by no means a summary all these approaches; instead we intend for it to be a careful meditation on two important concepts that will highlight the interplay between different mathematical tools in creating a fruitful theory of a multiresolution framework; the reader interested in the contributions of this work may peruse Section 1.4. We first revisit Mallat's original definition of MRAs; this will require some fundamental tools from Fourier Analysis which we introduce in the next section.

1.1 Fundamentals of Fourier Analysis

The theorems and techniques of Fourier Analysis are the cornerstones of the mathematics of signal and image processing. We begin with the notion of Hilbert space, the infinite dimensional analogue of $\mathbb{R}^n$ where we will conduct much of our work.

Definition 1.1.1. Let H be a vector space over $\mathbb{C}$ endowed with an inner product $\langle \cdot, \cdot \rangle : H \times H \mapsto \mathbb{C}$ such that $\langle x, x \rangle \geqslant 0$. Setting $\|x\| = \langle x, x \rangle$ we define the topology τ on H to be $\tau = \{T_{x,\delta}\}$ where $T_{x,\delta} = \{y \in H \mid ||x - y|| < \delta\}$. If the Cauchy sequences in H converge with

respect to this topology then H is called a *Hilbert space*. In other words, a Hilbert Space is a complete inner product space.

Remark 1.1.1. It should be noted that $\|\cdot\|$ satisfies the properties of a norm so τ is just the metric topology induced by the norm. However, a vector space B may not have an inner product and yet have a norm and be complete. In that case B is called a *Banach space*.

Example 1.1.1. Examples of Hilbert spaces:

1. $\mathbb{R}^n$
2. $l^2(\mathbb{C}) = \Big\{\{c_n\}_{n\in\mathbb{Z}} \mid c_n \in \mathbb{C}, \sum_{n\in\mathbb{Z}} |c_n|^2 < \infty\Big\}$. Here $l^2(\mathbb{C})$ is *the space of square summable sequences*. For the remainder of this thesis, we will refer to it as l^2 as we will only work with complex sequences.

Definition 1.1.2. Let $\Omega \subset \mathbb{R}^n$ and set $V_\mu^p(\Omega) = \{f : \Omega \mapsto \mathbb{C} \mid \big(\int_\Omega |f|^p \, d\mu\big)^{1/p} < \infty\}$ where μ is the Lebesgue measure on $\mathbb{R}$. Furthermore, define the equivalence relation $\sim$ on V_μ^p as $f \sim g \iff \int_\Omega |f|^p \, d\mu = \int_\Omega |g|^p \, d\mu$. Then the *space of p integrable functions*, termed L_μ^p is defined as $L_\mu^p = V_\mu^p / \sim$. We henceforth drop the subscript μ since we will work exclusively with the Lebesgue measure.

In the context of Fourier Analysis, we are particularly interested in the case where $p = 2$. The L^2 space, which in addition to being a Banach space, is also a Hilbert space. The formula for the inner product is a natural generalization of the formula for the 2-norm:

Definition 1.1.3. Let $f, g \in L^2(\Omega)$. The L^2 inner product between f and g is defined as

$$\langle f, g\rangle = \int_\Omega f\overline{g} \, dx. \tag{1.1.1}$$

Definition 1.1.4. Let $\Omega = (-\pi, \pi)$ and let $f \in L^2(\Omega)$. Then f can be written as a linear combination of complex exponentials:

$$f(x) \underset{a.e}{=} \sum_{n\in\mathbb{Z}} c_n e^{inx}, \tag{1.1.2}$$

where $c_n = (1/2\pi)\int_{-\pi}^{\pi} f(x)e^{inx} \, dx$. The series on the right of Equation 1.1.2 is called the *Fourier Series* of f and the coefficients c_n are called the *Fourier Coefficients*.

Remark 1.1.2. Using Euler's formula $e^{inx} = \cos(nx) + i\sin(nx)$ the Fourier Series of $f \in L^2(-\pi, \pi)$ may be written in the form

$$f(x) \underset{\text{a.e}}{=} a_0 + \sum_{n\in\mathbb{N}} a_n \cos(nx) + b_n \sin(nx). \tag{1.1.3}$$

Theorem 1.1.1 ([14, Proposition 3.2.7]). *Let $f \in L^2(-\pi, \pi)$ have Fourier coefficients $\{c_n\}_{n\in\mathbb{Z}}$. Then Parseval's identity states that*

$$\frac{1}{2\pi}\int_{[-\pi,\pi]} |f|^2\, dx = \sum_{n\in\mathbb{Z}} |c_n|^2. \tag{1.1.4}$$

Remark 1.1.3. Parseval's identity shows that the Fourier series is a norm preserving transformation from L^2 to l^2, and is thus an isometry.

The theory of Fourier series is just the beginning of the vast and expanding universe of Fourier Analysis. A key aspect of the latter is the extension of Fourier Series via *Pontryagin Duality*, to functions defined on $\mathbb{R}$ and on $\mathbb{Z}_n$.

Definition 1.1.5. Let $f \in L^1(\mathbb{R})$. The *Fourier Transform* and the *Inverse Fourier Transform* of $f \in L^1(\mathbb{R})$, denoted $\hat{f}$ and $\check{f}$ respectively are

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-i\xi x}\, dx \quad \text{and} \quad \check{f}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} f(\xi)e^{i\xi x}\, d\xi. \tag{1.1.5}$$

Theorem 1.1.2 ([12, Theorem 5.15]). *For $f \in L^1(\mathbb{R})$, $f(x) = \check{\hat{f}}(x)$ almost everywhere.*

Remark 1.1.4. Since $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, when $f \in L^2(\mathbb{R})$ we can set the Fourier transform of f to be the pointwise limit of the Fourier transforms of a sequence of L^1 functions converging to f in the L^2 sense. In other words, let $g_n \to f$ such that $g_n \in L^1(\mathbb{R})$. Then $\widehat{f}(\xi) := \lim_{n\to\infty} \widehat{g}_n(\xi)$.

Theorem 1.1.3 ([12, Lemma 5.19]). *Let $f, g \in L^2(\mathbb{R})$. Then*

$$\int_{\mathbb{R}} f(x)\overline{g(x)}\, dx = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}\, d\xi. \tag{1.1.6}$$

In particular, $\|f\|_2 = (1/\sqrt{2\pi})\|\widehat{f}\|_2$.

Remark 1.1.5. The function f is said to live in the time domain and its Fourier transform $\hat{f}$ is said to live in frequency domain because it tells us how much of each frequency, denoted ξ, is present in f.

Definition 1.1.6. Let $f : \mathbb{Z}_N \mapsto \mathbb{C}$. Then the *Discrete Fourier Transform* of f, denoted $\hat{f}$ is the function on $\mathbb{Z}_N$ given by:

$$\widehat{f}(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} f(k) e^{-i\frac{2\pi}{N}nk}. \tag{1.1.7}$$

1.2 Multiresolution Analysis

The idea behind a Fourier series is that the sequence of subspaces $D_n = \text{span}\{e^{inx}\}_{|n|\leqslant N}$ converges to $L^2([-\pi, \pi])$ as $N \to \infty$. In 1989, Mallat and Meyer [21] proposed multiresolution analyses to generalize this idea of approximating L^2 using sequences of subspaces.

Definition 1.2.1 ([21]). An *orthogonal multiresolution analysis* with a *scaling function* φ is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ satisfying the following properties:

1. Monotonicity:

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \cdots \tag{1.2.1}$$

   The subspace V_j is called a j-level approximation space.

2. Completeness:

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \tag{1.2.2}$$

3. Trivial intersection:

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\} \tag{1.2.3}$$

4. Scaling invariance: Let $f \in L^2(\mathbb{R})$. Then

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

5. Translation: $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for V_0.

The germs of Mallat and Meyer's idea were found in Alfred Haar's influential work *Zur theorie der orthogonalen funktionensysteme* where he proposed a basis for $L^2(\mathbb{R})$ using functions of compact support. His result provides an illustrative example of an MRA.

Example 1.2.1 ([16]). Let $\varphi(x) = \chi_{[0,1]}$ and define

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k), \quad j, k \in \mathbb{Z}.$$

We define the Haar approximation subspaces as follows:

$$V_0 = \text{span}\{\varphi(x-k)\}_{k\in\mathbb{Z}},$$
$$V_j = \text{span}\{\varphi_{j,k}\}_{k\in\mathbb{Z}}.$$

To prove that the Haar subspaces form a Multiresolution Analysis we verify each of the five properties from Definition 1.2.1.

1. Monotonicity

   Suppose $f \in V_j$ so $f = \sum_{k\in\mathbb{Z}} c_k\varphi(2^j x - k)$. Note that

   $$\varphi(2^j x - k) = \chi_{[\frac{k}{2^j}, \frac{k+1}{2^j}]} = \chi_{[\frac{2k}{2^{j+1}}, \frac{2k+2}{2^{j+1}}]} = \chi_{[\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}}]} + \chi_{[\frac{2k+1}{2^{j+1}}, \frac{2k+2}{2^{j+1}}]}.$$

   As a consequence,

   $$\begin{aligned} f(x) &= \sum_{k\in\mathbb{Z}} c_k\varphi(2^j x - k) \\ &= \sum_{k\in\mathbb{Z}} c_k\chi_{[\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}}]} + c_k\chi_{[\frac{2k+1}{2^{j+1}}, \frac{2k+2}{2^{j+1}}]} \\ &= \sum_{k\in\mathbb{Z}} c_k\varphi(2^{j+1}x - 2k) + c_k\varphi(2^{j+1}x - 2k - 1) \in V_{j+1}. \end{aligned}$$

2. Completeness

   We need to show that every square integrable function can be written as a (possibly infinite) linear combination of $\{\varphi_{j,k}\}_{j,k\in\mathbb{Z}}$. Let $f \in L^2(\mathbb{R})$. The idea is to approximate the Fourier transform of f with a compactly supported Fourier transform $\widehat{g}$ and then to approximate g through its projections on V_j. Note that we use the Fourier transform on f as the limit of Fourier transforms on a sequence of L^1 functions converging to f. Let $\varepsilon > 0$ be given and since $\int_{\mathbb{R}} |\widehat{f}|^2\, dx < \infty$ there exists $R > 0$ such that

   $$\int_{|x|\geqslant R} |\widehat{f}|^2\, dx < \varepsilon^2(\pi/2).$$

Let $\widehat{g} = \widehat{f}\chi_{[-R,R]}$. By Plancherel's theorem,

$$||f - g|| = \frac{1}{\sqrt{2\pi}}||\widehat{f} - \widehat{g}|| < \varepsilon/2. \tag{1.2.4}$$

Set $P_j(g) = \sum_{k\in\mathbb{Z}}\langle g, \varphi_{j,k}\rangle\varphi_{j,k}$. Note $P_j \in V_j$. The goal is now to calculate $||P_j(g)||_2$ and prove that it converges to $||g||_2$. Using the fact that $\widehat{\varphi_{j,k}} = 2^{-j/2}e^{-ik\xi/2^j}\widehat{\varphi}\big(\frac{\xi}{2^j}\big)$, by Plancherel's theorem we have

$$\begin{aligned}
|\langle g, \varphi_{j,k}\rangle|^2 &= \frac{1}{4\pi^2}|\langle \widehat{g}, \widehat{\varphi_{j,k}}\rangle|^2 \\
&= \frac{1}{4\pi^2}2^{-j}\Big|\int_{\mathbb{R}} \hat{g}(\xi)2^{-j/2}e^{-ik\xi/2^j}\overline{\widehat{\varphi}\big(\frac{\xi}{2^j}\big)}\,d\xi\Big|^2 \\
&= \frac{1}{4\pi^2}2^{j}\Big|\int_{\mathbb{R}} \hat{g}(2^j y)e^{-iky}\overline{\widehat{\varphi}(y)}\,dy\Big|^2 \\
&= \frac{1}{4\pi^2}2^{j}\Big|\sum_{m\in\mathbb{Z}}\int_{2m\pi}^{2(m+1)\pi} \hat{g}(2^j y)\overline{\widehat{\varphi}(y)}e^{-iky}\,dy\Big|^2 \\
&= 2^{j}\Big|\sum_{m\in\mathbb{Z}}\frac{1}{2\pi}\int_0^{2\pi} \hat{g}(2^j(t+2\pi m))\overline{\widehat{\varphi}(t+2\pi m)}e^{-ikt}\,dt\Big|^2 \\
&= 2^{j}\Big|\frac{1}{2\pi}\int_0^{2\pi}\sum_{m\in\mathbb{Z}} \hat{g}(2^j(t+2\pi m))\overline{\widehat{\varphi}(t+2\pi m)}e^{-ikt}\,dt\Big|^2.
\end{aligned}$$

The interchange of the sum and integral is allowed due to the uniform convergence of

$$H(t) := \sum_{m\in\mathbb{Z}} \hat{g}(2^j(t+2\pi m))\overline{\widehat{\varphi}(t+2\pi m)},$$

where $\widehat{g}$ has compact support so only finitely many terms are non-zero. Further notice that the value on the last line is just the kth Fourier coefficient of $H(t)$. Since the $\varphi_{j,k}$ are mutually orthogonal for each j (due to disjoint support) we have that

$$||P_j(g)||_2^2 = \sum_{k\in\mathbb{Z}}|\langle g, \varphi_{j,k}\rangle|^2 = 2^j\sum_{k\in\mathbb{Z}}\hat{H}(k) = \frac{2^j}{2\pi}\int_0^{2\pi}|H(t)|^2\,dt.$$

Here the last equality follows due to Parseval's identity for Fourier Series. To evaluate $\int_0^{2\pi}|H(t)|^2\,dt$, observe

$$|H(t)|^2 = \sum_{m\in\mathbb{Z}}\sum_{n\in\mathbb{Z}} \hat{g}(2^j(t+2\pi m))\overline{\widehat{\varphi}(t+2\pi m)}\hat{g}(2^j(t+2\pi n))\overline{\widehat{\varphi}(t+2\pi n)}.$$

We select a J such that for all $j > J$, we have no cross terms left in the summation. To that end, pick $2^J > R/\pi$, and suppose that for any $n \neq m$ there is a t such that

$\hat{g}(2^j(t+2\pi m))$ and $\hat{g}(2^j(t+2\pi n))$ are simultaneously non-zero. Then since $\widehat{g}(\xi)=0$ for $|\xi|>R$, we must have that $2^j(t+2\pi m)<R$ and $2^j(t+2\pi n)<R$. Then

$$|2\pi| \leqslant |2\pi(m-n)| < |2\pi m + t| + |-t-2\pi n| \leqslant \frac{R}{2^j} + \frac{R}{2^j} < 2\pi.$$

Contradiction. So for every t, $\hat{g}(2^j(t+2\pi n))\hat{g}(2^j(t+2\pi m)) \neq 0$ only when $m=n$. As a consequence,

$$|H(t)|^2 = \sum_{m\in\mathbb{Z}} |\hat{g}(2^j(t+2\pi m))\widehat{\varphi}(t+2\pi m)|^2.$$

Employing this to calculate $||P_j(g)||_2^2$, we have

$$\begin{aligned}
||P_j(g)||_2^2 &= \frac{2^j}{2\pi}\int_0^{2\pi} |H(t)|^2\,dt \\
&= \frac{2^j}{2\pi}\int_0^{2\pi} \sum_{m\in\mathbb{Z}} |\hat{g}(2^j(t+2\pi m))\widehat{\varphi}(t+2\pi m)|^2\,dt \\
&= \frac{2^j}{2\pi}\sum_{m\in\mathbb{Z}}\int_0^{2\pi} |\hat{g}(2^j(t+2\pi m))\widehat{\varphi}(t+2\pi m)|^2\,dt \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(\xi)|^2 |\widehat{\varphi}\Big(\frac{\xi}{2^j}\Big)|^2\,d\xi.
\end{aligned}$$

Next we want to take the limit of the right side with respect to j but must slide it through the integral somehow. The Dominated Convergence Theorem provides a way to do that because

$$|\widehat{g}(\xi)|^2 |\widehat{\varphi}\Big(\frac{\xi}{2^j}\Big)|^2 \leqslant ||\widehat{\varphi}||_\infty |\widehat{g}(\xi)|^2 = d.$$

So we have a dominating function d and $||d||_1 = \int_{\mathbb{R}} ||\widehat{\varphi}||_\infty |\widehat{g}(\xi)|^2 < \infty$ because $\widehat{g}\in L^2(\mathbb{R})$. Thus

$$\lim_{j\to\infty} ||P_j(g)||_2^2 = \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(\xi)|^2 \lim_{j\to\infty} |\widehat{\varphi}\Big(\frac{\xi}{2^j}\Big)|^2\,d\xi = \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(\xi)|^2\,d\xi = \frac{1}{2\pi}||\widehat{g}||_2^2 = ||g||_2^2.$$

As a consequence, there exists $K>J$ such that for all $j>K$, $||P_j(g)-g||_2 < \varepsilon/2$. Finally, combining this estimate with Equation 1.2.4 we have that

$$||f-P_j(g)||_2 = ||f-g+g-P_j(g)||_2 \leqslant ||f-g||_2 + ||P_j(g)-g||_2 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

3. Trivial intersection

   Suppose $f \in \bigcap_{j\in\mathbb{Z}} V_j$. Then f must be constant on every dyadic interval; in particular it is constant on $[0, 2^j]$ for every $j \in \mathbb{Z}$. So $f(x) = c \in \mathbb{C}$ for $x \in \mathbb{R}$, so $c = 0$ since $f \in L^2(\mathbb{R})$.

4. Scaling invariance

   Let $f \in V_j$ which means $f = \sum_{k\in\mathbb{Z}} a_k \varphi(2^j x - k)$. So $f(2x) = \sum_{k\in\mathbb{Z}} a_k \varphi(2^{j+1}x - k)$. The converse is identical.

5. The translation property holds trivially based on the definition of V_0.

1.2.1 Wavelet Spaces

Since $V_j \subset V_{j+1}$, we can find the orthogonal subspace W_j of V_j in V_{j+1}. Consequently, we can write $V_{j+1} = V_j \bigoplus W_j$. Using successive decompositions of V_j we have

$$V_{j+1} = \bigoplus_{n\leqslant j} W_n.$$

Finally, since $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$, we get

$$L^2(\mathbb{R}) = \bigoplus_{n\in\mathbb{Z}} W_n \tag{1.2.5}$$

The W_j are the *wavelet* or *detail* spaces as they capture the details in going from a level j resolution V_j to a level $j+1$ resolution V_{j+1}. For the Haar MRA, we can write down an explicit basis for W_j. The function $\psi(x) = \chi_{[0,1/2]} - \chi_{[1/2,1]}$ is known as the *Haar (Mother) Wavelet*, and the Haar Wavelet spaces are given by dyadic translations and dilations of ψ.

$$W_j = \text{span}\{2^{j/2}\psi(2^j x - k)\}_{k\in\mathbb{Z}}.$$

At each level j the corresponding wavelet $\psi_{j,k}$ is supported on the interval $[\frac{k}{2^j}, \frac{(k+1)}{2^j}]$. Consequently, the dilation factor determines the size of the support and the translation factor determines its location. Moreover, at a fixed level, the supports are disjoint. These properties are useful for expanding a given function in the Haar Basis.

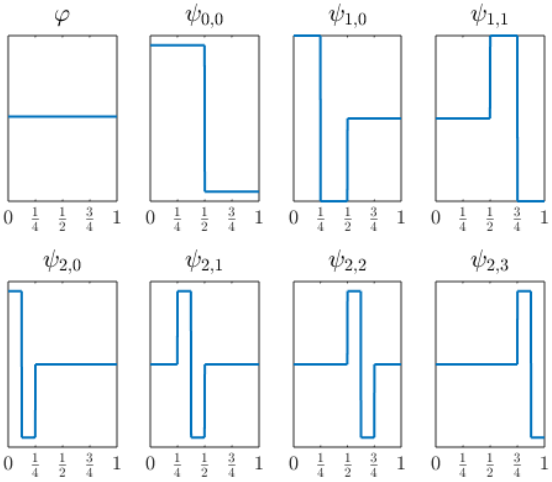


Figure 1.1: Haar wavelets $\{\psi_{j,k}\}_{j=0,1,2}$ on [0,1] with scaling function $\varphi = \chi_{[0,1]}$. Each $\psi_{j,k}$ is a square wave supported on an interval that gets smaller as j increases.

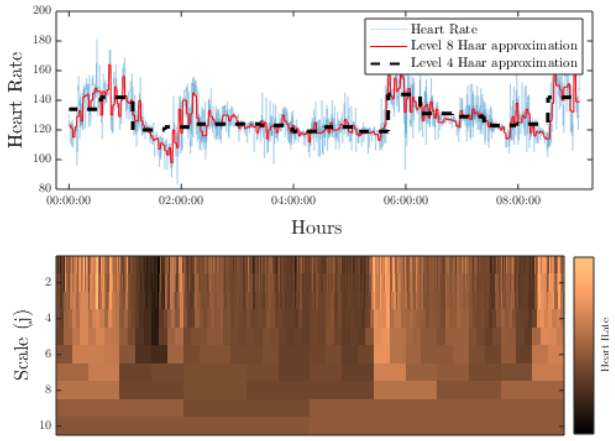


Figure 1.2: Top: We plot heart rate data (light blue) and its level 4 and 8 Haar approximations. The level 4 and 8 approximations are constant on disjoint dyadic intervals of size 2^{-4} and 2^{-8} respectively. Bottom: We plot all the Haar approximations where the heart rate is given by colour. Notice that the approximations improve by partitioning $[0, 1]$ by half each time.

In Figure 1.2 the level 4 Haar approximation picks up broader variations in the heart rate signal while level 8 picks up fine variations so at increasingly higher scales, we create more detailed approximations to the original signal. This is exactly Mallat's point: we *don't*

need to alter the sampling of f. Instead, approximating it in the right basis automatically introduces the desired resolution. The Haar Wavelet encodes *resolution* through factor j which partitions $[0,1]$ into dyadic intervals of size 2^{-j} by the oscillations of $\psi_{j,k}$. At a fixed level j the Haar basis function $\psi(2^j x - k)$ picks up variations of f in the dyadic interval indexed by k. In this way, the Haar decomposition is *adaptive* or *localized* as it adapts to the function's behaviour in different regions of $[0,1]$. Lastly, the Haar wavelet forms an orthonormal basis for $L^2[0,1]$ so each L^2 function can be written as a linear combination in the Haar basis, with coefficients recovered using the projection formula. Consequently, the Haar wavelet enables *efficient reconstruction*. These three properties make the Haar wavelet an attractive model for decomposing one-dimensional signals, and the definition of the MRA describes these three desirable properties mathematically.

Example 1.2.2 (2-D Haar wavelet). An image is understood as a function $g : [0,1]^2 \mapsto \mathbb{R}$ where $g(x)$ is the pixel value at x. Just as we built a Haar basis by dyadic partitions of the interval, we may build a Haar basis through dyadic partitions of the unit square and then express an image in the Haar basis. The result shown in Figure 1.3 shows how 2-D Haar wavelets can compress, or "pixellate", an image through lower order approximations. The details of the 2-D implementation of the Haar Wavelet are found in [31].

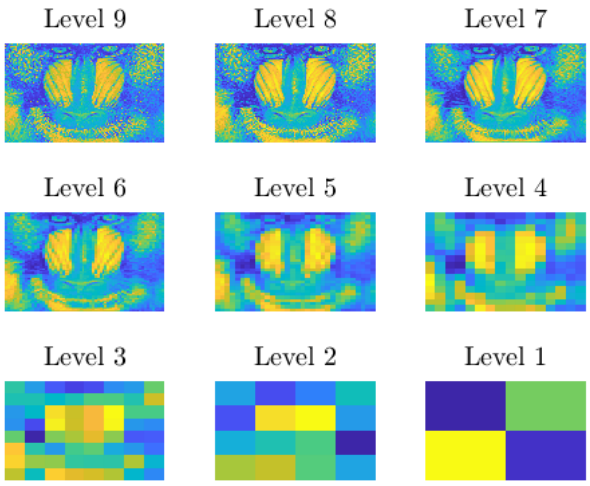


Figure 1.3: Successive 2-D Haar approximations to the image of a Mandrill

Example 1.2.3 (Littlewood-Paley form of a Fourier Series [24]). The main point of classical Fourier analysis is that any square integrable periodic function is equal to its Fourier series in the L^2 norm. In other words, $\lim_{N\to\infty} ||f - T_N f||_2^2 = 0$, where $T_N f = \sum_{|n|\leqslant N} c_n e^{inx}$. However, convergence in L^2 does not imply pointwise convergence; Lusin actually conjectured that the convergence for the exponentials in particular was pointwise almost everywhere [28]. Carleson showed this conjecture to be true [5]; but on the other hand, Kolmogorov constructed an L^1 function whose Fourier Series converged nowhere but a countable set of points [30]. Thus, there is something special about using the exponential functions as a basis for $L^2([-\pi,\pi])$. Naturally, we must ask if we can use the exponentials to form a multiresolution analysis. The answer is: almost! We first consider the case of approximation spaces on $[-\pi,\pi]$. Let the scaling function be $\phi = \chi_{[-\pi,\pi]}$ and define V_j as

$$V_j = \{f \in L^2([-\pi,\pi]) \mid f = \sum_{|n|\leqslant 2^j} c_n e^{inx}\}.$$

Let us now check if the properties for an MRA hold, and the extent to which they fail.

1. Monotonicity: Any trigonometric polynomial up to frequency 2^j can be written as a trigonometric polynomial up to 2^{j+1} by setting the Fourier coefficients c_i zero for $2^j < i \leqslant 2^{j+1}$. The constant function ϕ is thus in every subspace V_j and the monotonicity condition changes to one-sided monotonicity: $V_0 \subset V_1 \subset V_2 \subset \cdots$.

2. Scaling: To see sufficiency, let $f \in V_j$. Then $f = \sum_{|n|\leqslant 2^j} c_n e^{inx}$ and $f(2x) = \sum_{|n|\leqslant 2^j} c_n e^{i2nx} = \sum_{|n|\leqslant 2^{j+1}} d_n e^{inx}$ where $d_{2n} = c_n$ and 0 otherwise. For necessity, note that if $f(2x) \in V_{j+1}$ then

$$\begin{aligned} f(2x) &= \sum_{n=1}^{2^{j+1}} d_n e^{inx} = \sum_{k=1}^{2^j} d_{2k} e^{ik2x} + \sum_{k=0}^{2^j-1} d_{2k+1} e^{i(2k+1)x} \\ &= \sum_{k=1}^{2^j} d_{2k} e^{ik2x} + e^{ix} \sum_{k=0}^{2^j-1} d_{2k+1} e^{ik2x} = \sum_{k=1}^{2^j} d_{2k} e^{ik2x} + e^{i/2} \sum_{k=0}^{2^j-1} d_{2k+1} e^{i(k+1)2x}. \end{aligned}$$

   Using the transformation $y = 2x$ we get that

$$f(y) = \sum_{k=1}^{2^j} d_{2k} e^{iky} + e^{i/2} \sum_{k=0}^{2^j-1} d_{2k+1} e^{i(k+1)y} \in V_j.$$

3. Completeness: This is a consequnce of completeness of the exponential basis because

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = \overline{\{f \in L^2([-\pi,\pi]) \mid f = \sum_{|n|\leqslant 2^j} c_n e^{inx} \text{ for some } j \in \mathbb{N}\}} = L^2([-\pi,\pi]).$$

4. The translation property doesn't apply here since ϕ already spans the whole space V_0.

5. Due to one-way monotonicity, the intersection of these spaces is not trivial: constant functions are contained in every approximation space.

In Example 1.2.1 we mentioned that an ideal MRA should provide resolution, be localized, and exhibit efficient reconstruction. The modes of complex exponentials provide a natural resolution for $[-\pi,\pi]$ because as n increases, the real and imaginary parts of e^{inx} oscillate more rapidly, thus capturing the notion of detail via partition seen in the Haar wavelet. In this way, n in e^{inx} plays a similar rule to j in $\psi_{j,k}$. Complex exponentials, like the Haar wavelet, also form an orthonormal basis for $L^2([-\pi,\pi])$ so the Fourier coefficients can be easily recovered using the projection formula. However, the Fourier basis is not localized because sines and cosines do not have compact support like the Haar wavelet.

1.3 The Paradigm of Graphs

Despite their wide-ranging applicability, MRAs are sometimes inefficient at solving modern problems in image processing. For example, consider the following question. Given the image of Barack Obama in Figure 1.4, can we isolate the pixels that indicate his face?

Figure 1.4: A picture of former U.S. President Barack Obama

The Haar decomposition of the image can successfully identify contours, but a more efficient solution emerges when we think of the image as a weighted graph. There are several approaches to associate a graph with an image; these are well summarized in [23]. Here we present one simple method that suffices to solve the above problem. Given an image I as a vector of pixel location and luminescence we form the *graph associated with the image* Γ_I as follows: Each vertex of Γ_I represents a pixel. To assign weights to edges, we propose that two pixels are similar to each other if they are close to each other and they have similar luminescence. Consequently, given two pixels i and j we assign the edge weight

$$w(i,j) = \exp\left(\frac{-||x_i - x_j||_2^2}{2\sigma_x}\right) \exp\left(\frac{-||p_i - p_j||_2^2}{2\sigma_p}\right)$$

Here x_i and x_j are the positions of the pixels in a lattice contained within $[0,1]^2$, p_i and p_j are the luminescences at i and j, and σ_x and σ_p are position and luminescence sensitivities respectively. Since we have a weighted graph Γ_I we use the tools of Spectral Graph theory explained in Section 2.2. We compute the spectrum of the Laplacian of Γ_I and classify vertices based on the parity of the eigenvector corresponding to the minimal non-zero eigenvalue in Figure 1.5. Somehow f is able to almost precisely tell us the pixels that correspond to the face! The reason is explained in Section 2.2.

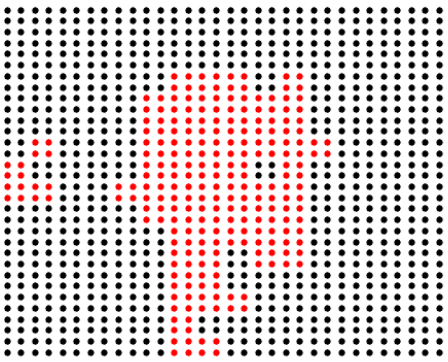


Figure 1.5: Positive (black) and negative (red) values of f: each dot (or vertex) represents a pixel and the red coloured dots identify the face. Note that we were able to identify the face only using information on pixel luminescence and location.

1.4 Thesis Contribution

In Chapter 2 we revisit the curious failure of the Fourier basis as an ideal MRA on $[-\pi, \pi]$ and examine whether the Graph Fourier Basis as the orthogonal eigenbasis of the Graph Laplacian improves on these failures. We show–in the spirit of the Sturm Comparison Principle–that the resolution property holds in a weaker sense on graphs (Proposition 2.2.1). Secondly, we introduce the localization phenomenon in Graph Fourier Bases and prove (Propositions 2.3.1 and 2.3.4) that it is equivalent to the existence of joint Dirichlet-Neumann eigenfunctions, which do not exist on Euclidean domains. In Chapter 3, we introduce Belkin and Niyogi's Diffusion Maps as a loose Multiresolution Analysis on graphs which can be used to draw and detect low-dimensional manifolds encoded in high dimensional data. In Section 3.3 we propose and implement Neumann maps to improve Diffusion Maps in terms of embedding submanifolds within manifolds. All the results mentioned in this section are novel; furthermore, the construction of Neumann maps is an application of the symmetrization idea in [7] to the reflecting random walk found in [6, Chapter 8] but has not (to the author's knowledge) appeared in previously published work on the subject. Lastly, all the simulations and figures have been written by the author in `MATLAB` with the aid of the toolbox in [25]. The code can be found in the Appendix at the end of this work.

Chapter 2

Fourier Analysis on Graphs

Example 1.2.3 shows that the Fourier Basis of exponentials $\{e^{inx}\}_{n\in\mathbb{Z}}$ forms a partial MRA for $L^2([-\pi,\pi])$. The drawback of the Fourier basis is that the basis functions do not exhibit the compact support that the Haar wavelets $\psi_{j,k}$ do which makes it a poor candidate as an image compression scheme because it is unable to adapt to different regions of the image. As an illustrative example, consider the Haar and Fourier coefficients of the 32 bit image in Figure 2.1.

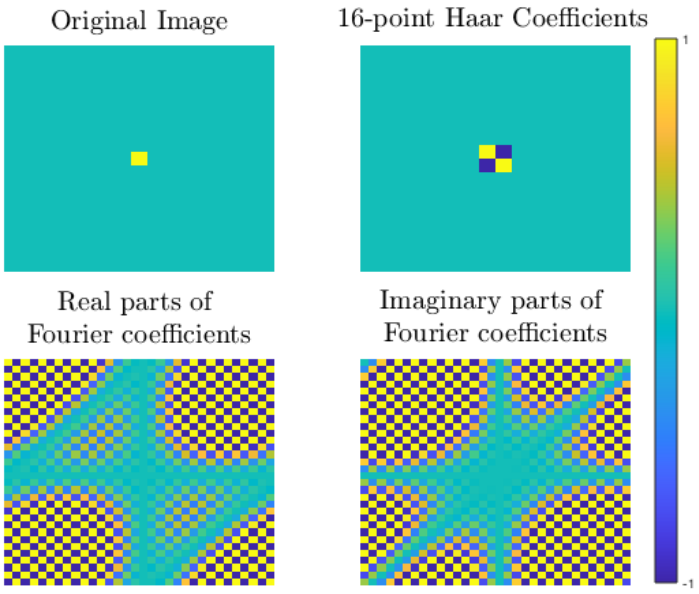


Figure 2.1: Comparing the 2-D Haar and DFT coefficients

There are only 4 non-zero Haar coefficients but several small non-zero Fourier coeffi-

cients so the image is more conveniently stored in terms of Haar coefficients than Fourier coefficients. This is due to the following relationship between the smoothness a function and the decay of its Fourier coefficients:

Theorem. *Let* $f \in L^2([-\pi, \pi])$ *and* $f \underset{a.e}{=} \sum_{n\in\mathbb{Z}} c_n e^{inx}$. *Then if* $f \in C^k$ *then* $c_n = O(n^{-k})$.

An image with a blob is like a C^0 function that is not C^1 so its Fourier coefficients decay slowly like $1/n$. That is why we see several small but non-zero coefficients in the turquoise regions of the real and imaginary parts of the Fourier Coefficients.

In this section, we will present a theory for Fourier Series on graphs which will hopefully fix problems such as non-localization seen in Fourier Series on intervals. First we will use the spectral theorem for the Euclidean Laplacian to prove why $\{e^{inx}\}$ emerges as a basis of functions for $L^2([-\pi, \pi])$. Then using the spectral theorem for the graph Laplacian we will present an analogue of a Fourier basis for graphs. In the latter part of the section, we will compare the Fourier bases in the two settings. In particular, we find that the resolution property of the Fourier basis on $[-\pi, \pi]$ holds in a weaker sense. On the other hand, we present some graphs where the Fourier basis is localized, and hence adaptive. We prove that this phenomenon on graphs is equivalent to the existence of joint Dirichlet-Neumann spectra on subgraphs, which is again not possible in the Euclidean case. However, the localized basis functions are sometimes not orthogonal so we compromise on the efficient reconstruction property of the Fourier Series.

2.1 Properties of eigenfunctions of Δ on $[-\pi, \pi]$

In this section we justify the important properties of the Fourier basis $\{e^{inx}\}_{n\in\mathbb{Z}}$ from the point of view of spectral theory. Recall that in Chapter 1 we demonstrated the following facts about the complex exponentials (or equivalently trigonometric functions):

1. **Resolution.** Let S_n (resp. C_n) be the set of vanishing points of $\sin(nx)$ (resp $\cos(nx)$). Then $|S_n| = 2n + 1 = |C_n| + 1$. The control on the number of oscillations as a function of frequency n allows the notion of "fineness" to be encoded within the trigonometric functions.

2. **Non-adaptivity.** $\sin(nx)$ and $\cos(nx)$ do not vanish on any open subsets of $[-\pi, \pi]$.

3. **Reconstruction and completeness.** Let $f : [-\pi, \pi] \mapsto \mathbb{R}$ and $\int_{[-\pi,\pi]} |f|^2 \, dx < \infty$. Then

$$f(x) = \sum_{n \in \mathbb{Z}} c_n e^{inx} \text{ where } c_n = \int_{[-\pi,\pi]} f(x) e^{inx} \, dx.$$

We demonstrate that properties (2) and (3) also hold for the set of Laplacian eigenfunctions on a Euclidean domain $\Omega \subset \mathbb{R}^n$. Furthermore, property (1) is a special property of Laplacian eigenfunctions on 1-Dimensional connected manifolds. Before we prove properties (1) and (2), we state all the relevant definitions that will lead to a statement of the spectral theorem for compact self-adjoint operators on separable Hilbert spaces that will justify property (3). The discussion in this section is a curation of the spectral theory of linear elliptic equations presented in [11] and [13]. For more detailed intuition, refer to the Appendix. Furthermore, we denote the space of k-times differentiable functions as $C^k(\Omega)$ and the space of compactly supported k-times differentiable functions as $C_c^k(\Omega)$. We will occasionally drop Ω when there is no ambiguity about the domain.

Definition 2.1.1. Let $u : \Omega \mapsto \mathbb{R}$, $u \in C^2$. The *Laplacian of* u, denoted Δu, is the function

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}.$$

Definition 2.1.2. We say that $\lambda \in \mathbb{R}$ and $u \in C^2(\Omega)$ are, respectively, a *Dirichlet eigenvalue* and a *Dirichlet eigenfunction* of the Laplacian Δ if they satisfy

$$\begin{aligned} -\Delta u + \lambda u &= 0 \quad \text{in } \Omega, \\ u &\equiv 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.1.1}$$

When equation 2.1.1 is replaced by $\partial_n u = 0$ then u and λ are called *Neumann* eigenfunction and eigenvalue respectively.

The main lesson of the spectral theory of elliptic operators is that $C^2(\Omega)$ is not the right setting to study solutions to the equation $\Delta u + \lambda u = f$ where $f \in L^2(\Omega)$. Instead, it turns out that the theory is much richer when we switch to the setting of weak derivatives. For the remainder of this section, we assume $\Omega \subseteq \mathbb{R}^n$ is an open simply connected domain.

Definition 2.1.3. Let $u, v_i \in L^1$ for $1 \leqslant i \leqslant n$ and set $g_i(x) = u(x_1, \ldots, x, x_{i+1}, \ldots, x_n)$ as the restriction of u to the ith coordinate and suppose they satisfy

$$\int_{\Omega_i} g_i \varphi \, dx = -\int_{\Omega_i} v_i \varphi \, dx.$$

for every $(x_1, \ldots, x_n) \in \Omega$ and $\varphi \in C_c^\infty(\Omega_i)$ where Ω_i is the projection of Ω in the ith coordinate. Then v_i is a *weak (partial) derivative* of u in the ith direction, denoted $v = u_i$. If u has weak derivatives in every direction then the *weak derivative of* u is the vector of weak partial derivatives of u denoted ∇u.

Definition 2.1.4. We may keep taking weak partial derivatives in all possible directions and define the *Sobolev Space* $W^{k,p}$ as the space of functions whose first k weak derivatives are in L^p, i.e.,

$$W^{k,p}(\Omega) = \{u : u \text{ has } k \text{ weak derivatives and } ||\nabla^j u||_p < \infty \text{ for } 1 \leqslant j \leqslant k\},$$

where $||\nabla^j u||_p^p = \displaystyle\int_\Omega \sum_{m \subset [n], |m| = j} |u_m|^p \, dx$. We also define the space

$$W_0^{k,p}(\Omega) = \overline{C_c^\infty \cap W^{k,p}}.$$

Here, the closure is with respect to the Sobolev norm $||u||_{W^{k,p}} = \sum_{1 \leqslant j \leqslant k} ||\nabla^j u||_p$. As usual, in the case $p = 2$, Sobolev spaces are also Hilbert spaces (since their weak derivatives are in L^2) so we denote $W^{1,2} = H^1$ and $W_0^{1,2} = H_0^1$. Note, however, that the H^1 and L^2 inner products are different

$$\langle f, g \rangle_{L^2} = \int_\Omega f \overline{g} \, dx \quad \text{versus} \quad \langle f, g \rangle_{H^1} = \int_\Omega f \overline{g} + \nabla f \cdot \overline{\nabla g} \, dx.$$

The main point of introducing Sobolev spaces is to state the weak formulation of the eigenvalue problem 2.1.2.

Definition 2.1.5. Let $u \in H_0^1$, $f \in L^2$, and $\lambda \in \mathbb{R}$ satisfy

$$\int_\Omega \nabla u \cdot \nabla v + \lambda u v \, dx = \int_\Omega f v \, dx \quad \text{for every } v \in H_0^1. \tag{2.1.2}$$

When $f = 0$, u is a *weak Dirichlet eigenfunction.* Similarly, if H_0^1 is replaced with H^1 then u is a *weak Neumann eigenfunction.*

The following proposition defines Δ in the "weak sense" on H_0^1.

Proposition 2.1.1. *Let $-\Delta u + \lambda u = f$ if and only if there exists $f \in L^2(\Omega)$ such that Equation 2.1.2 is satisfied. Then Δ is well-defined as an operator from H_0^1 to L^2.*

From here on, we will think of Δ in the sense of Proposition 2.1.1 with the understanding that when u is sufficiently smooth, it is the same as the operator in Definition 2.2.5. The next proposition justifies the terminology of Dirichlet and Neumann in Definition 2.1.5:

Proposition 2.1.2. *Let $u \in C^2$ be a Dirichlet (resp. Neumann) eigenfunction. Then u is a weak Dirichlet (resp. Neumann) eigenfunction.*

The main point of working in Sobolev spaces is that the converse to Proposition 2.1.2 holds. This is a consequence of the **Elliptic Regularity Theorem** or the stronger **Analytic Regularity Theorem**. To state the analytic regularity theorem we recall the definition of multivariate analyticity.

Definition 2.1.6. Let $f : U \mapsto \mathbb{R}$ where $U \subseteq \mathbb{R}^n$. Then f is analytic if for every $x_0 \in U \; \exists \, r_{x_0}$ s.t. for $|x - x_0| \leqslant r_{x_0}$

$$f(x) = \sum_{t \in \mathbb{N}_0^n} a_t (x - x_0)^t,$$

with the power series on the right being absolutely convergent in the given neighbourhood. Here $\mathbb{N}_0^n$ is the set of all n-tuples of the natural numbers including 0 and $x^t = \Pi_{i=1}^n x_i^{t_i}$.

Theorem 2.1.3 (Analytic Regularity Theorem [13, Theorem 29]). *If $u \in H_0^1(\Omega)$ such that $-\Delta u + \lambda u = f$ and f is analytic then u is also analytic in Ω.*

The uniqueness principle for analytic functions provides the following result on the non-localization of Laplacian eigenfunctions:

Corollary 2.1.4. *Dirichlet (resp. Neumann) eigenfunctions cannot vanish in any open neighbourhood of Ω.*

Thus we see that weak Dirichlet (resp. Neumann) Eigenfunctions, when they exist, are analytic and solve the Dirichlet (resp. Neumann) problem. But why should they exist in the first place? And why should their completion be all of $L^2(\Omega)$? This is where we switch to the theory of operators on Hilbert Spaces.

Definition 2.1.7. Let $T : X \mapsto Y$ be a linear operator between Hilbert spaces over $\mathbb{C}$ with the metric topology induced by the norm. Then

1. T is compact $\iff$ for every bounded set $B \subset X$, $\overline{T(B)}$ is compact in Y $\iff$ every bounded sequence f_n has a convergent subsequence in Tf_n.
2. T is self-adjoint $\iff \langle x, Ty\rangle = \langle Tx, y\rangle$.
3. T is positive $\iff \langle x, Tx\rangle > 0$.
4. T is bounded $\iff \exists\, M$ such that $||Tu||_Y \leqslant M||u||_X$.

Example 2.1.1. Δ on $H_0^1(\Omega)$ is bounded and self-adjoint, but not compact or positive.

1. Δ is self-adjoint: Let $u, v \in H_0^1$. Then according to the weak formulation of $-\Delta u = f$ we have that $\langle -\Delta u, v\rangle = \langle f, v\rangle = \int_\Omega \nabla u \cdot \nabla v\, dx = \langle u, g\rangle = \langle u, -\Delta v\rangle$ where $-\Delta v = g$. The same proof holds in H^1.
2. Δ is not compact: Let $\Omega = [-\pi, \pi]$ and $f_n = \frac{1}{2\pi} e^{inx}$. Then $\Delta f_n = \frac{-n^2}{2\pi} e^{inx}$. However, $||\Delta f_n - \Delta f_m||_2 = \sqrt{n^4 + m^4} \geqslant \sqrt{2}$. Thus, $\{\Delta f_n\}$ is not Cauchy so since $L^2([-\pi, \pi])$ is complete, it is not convergent.
3. $\Delta : H_0^1(\Omega) \mapsto L^2(\Omega)$ is bounded as an operator. This is because of the way the Riesz Representation Theorem used in Proposition 2.1.1 gives the L^2 norm of $\Delta u = f$

$$||\Delta u||_2 = \sup_{v \in H_0^1} \frac{\langle f, v\rangle}{||v||_{H^1}} = \sup_{v \in H_0^1} \frac{\int_\Omega \nabla u \cdot \nabla v\, dx}{||v||_{H^1}} \leqslant ||u||_{H^1}.$$

4. Δ is not positive: The counterexample occurs when $\Omega = [-\pi, \pi]$ and $u = \sin(x)$.

We are now ready to state the spectral theorem:

Theorem 2.1.5 (Spectral Theorem [13, Theorem 6]). *Let $A : H \mapsto H$ be a compact, self-adjoint, positive operator on the Hilbert space H. When H is separable, i.e it admits a countable basis, H admits a countable orthogonal basis $\{u_n\}_{n\in\mathbb{Z}}$ and a decreasing sequence $\lambda_n \geqslant 0$ such that*

1. *$Ku_n = \lambda_n u_n$,*

2. $\lambda_n = \sup_{u \in H_{n-1}^{\perp}} \frac{\langle u, Ku \rangle}{||u||^2} = S_n(u)$. *Here* $H_{n-1} = span(u_1, \ldots, u_{n-1})$.

The spectral theorem states that a compact, self-adjoint, and positive operator A admits a *countable spectrum* and that the eigenfunctions span the Hilbert space when it is separable. Additionally, the eigenvalues can be computed variationally through the quotient found in part (2). However, we must note that u_i may not always be the minimizing element of the functional $S_n(u)$. Furthermore, the spectral theorem cannot be directly applied to Δ on $L^2(\Omega)$ because Δ is not positive or compact. However, the *Resolvent* [13] gives an equivalent formulation of the eigenvalue problem and helps us apply Theorem 2.1.5:

Definition 2.1.8. Fix $t \in \mathbb{C}$. Then the *resolvent* R_t is the operator:

$$R_t = (\Delta - tI)^{-1}.$$

Remark 2.1.1. It is a priori unclear whether $\Delta - tI$ is even invertible for any t. However, the following lemma shows that such a t must exist because of the boundedness of Δ.

Lemma 2.1.6. *R_t is well-defined for some $t \in \mathbb{C}$.*

Proof. First we note that for an operator between Hilbert spaces $A : H \mapsto K$, $(I - A)$ is invertible when $||A|| < 1$ and the inverse is given by

$$(I - A)^{-1} = \sum_{n \geqslant 0} A^n.$$

This is a "geometric series" for operators whose convergence is shown by proving that partial sums are Cauchy using the triangle inequality and the $||A|| < 1$ condition. The only preliminary step is to note that the vector space of bounded operators on H to K is itself complete under the metric topology induced by the norm. Invertibility follows from multiplying the series on the right by $I - A$, therefore making it to telescope to I. Now note that

$$(\Delta - tI) = -t(I - \Delta/t).$$

The operator on the right is invertible whenever $|t| > ||\Delta||$ due to the geometric series. □

Next, we prove that the resolvent has all the necessary properties as an operator from $L^2(\Omega)$ to $L^2(\Omega)$.

Lemma 2.1.7. *R_t is self-adjoint, positive, and compact.*

Proof. Self-adjointedness follows from the Neumann series expression of R_t. R_t is positive by the triangle inequality. To prove compactness, note that $\Delta + tI$ acts from $H_0^1(\Omega)$ to $L^2(\Omega)$ so $R_t = (\Delta + tI)^{-1}$ acts on $L^2(\Omega)$ to H_0^1. Furthermore, $\Delta + tI$ is bounded and bijective (since it has a well-defined inverse) so by the open mapping theorem its inverse is bounded. Hence R_t is bounded in H_0^1. Now let $\{x_n\}$ be a bounded sequence in L^2. Then $R_t x_n$ is bounded in H_0^1. But by the Rellich-Kondrachov theorem [11], a bounded sequence in H_0^1 has a convergent subsequence in $L^2(\Omega)$. Thus, we have that $R_t : L^2 \mapsto L^2$ is a compact operator. □

Lemma 2.1.8. *Let $u \in H_0^1(\Omega)$. Then $\Delta u = \lambda u$ (in the weak sense) $\iff$ $R_t u = \frac{1}{t+\lambda} u$ (in the weak sense).*

Proof. $\Delta u = \lambda u \iff \Delta u + tu = \lambda u + tu = (\Delta + tI)u = (t+\lambda)u \iff \frac{1}{t+\lambda} u = (\Delta + tI)^{-1} u = R_t u$. □

Theorem 2.1.9. *The solutions to the Dirichlet eigenvalue problem $\{u_n\}_{n\in\mathbb{N}}$ form a countable basis for $D(\Omega)$. Furthermore, the eigenvalues are given by*

$$\lambda_n = \inf_{u \in H_{n-1}^{\perp}} \frac{\langle u, Ku \rangle}{||u||_2^2} = \inf_{u \in H_{n-1}^{\perp}} \frac{\int_\Omega |\nabla u|^2 \, dx}{\int_\Omega u^2 \, dx} = S_n(u). \tag{2.1.3}$$

Here $H_{n-1} = span(u_1, \ldots, u_{n-1})$ and S_n is the Rayleigh quotient.

Proof. First, note that from Example 1.2.1 we have that $D(\Omega) \subset L^2(\Omega)$ is separable because the Haar wavelets $\{\psi_{n,k}\}_{n,k\in\mathbb{Z}}$ form a countable orthogonal basis for $L^2(\Omega)$. Moreover, due to Lemma 2.1.6 R_t exists and Lemma 2.1.7 combined with Theorem 2.1.5 proves that the eigenfunctions of R_t form a countable basis for D. However, Lemma 2.1.8 proves that the eigenfunctions of R_t are the Dirichlet eigenfunctions of Δ so the eigenfunctions of Δ form a countable orthogonal basis for D. Equation 2.1.3 is a consequence of Part (2) of Theorem 2.1.5 and Lemma 2.1.8. □

Remark 2.1.2. The above argument is also true for functions with Neumann conditions. Most of the proof is identical, except for that of the compactness of the resolvent because

it relies on the embedding of $H_0^1(\Omega)$ in $L^2(\Omega)$. Neumann eigenfunctions are not necessarily in $H_0^1(\Omega)$ so we cannot use this result. However, if we assume that the boundary of Ω is Lipschitz then the resolvent is compact.

Lemma 2.1.10. *Let $\Omega \in \mathbb{R}^n$ be a bounded domain with smooth boundary. Then if $\exists u$ such that $\Delta u = 0$ and $u = 0$ on $\partial\Omega$ then $u \equiv 0$.*

Proof. Note that due to Corollary 2.1.3, we have that $u \in C^\infty$ so we can freely use integration by parts. Consequently, $\int_\Omega |\nabla u|^2\, dx = \int_\Omega \nabla u \cdot \nabla u\, dx = \int_\Omega \nabla \cdot (u\nabla u)\, dx - \int_\Omega u\Delta u = 0$ due to the divergence theorem and the harmonic condition. Thus, $u_{x_i} = 0$ for every $i \leqslant n$ so $u \equiv C$ on Ω. The continuity of u up to $\partial\Omega$ proves that $C = 0$. □

Theorem 2.1.11. *Let $\Delta u = \lambda u$ for some $\lambda \in \mathbb{R}$ such that $u = \nabla u \cdot \hat{n} = 0$ on $\partial\Omega$. Then $u \equiv 0$.*

Proof. We use Rellich's identity [26] for Dirichlet eigenvalues which states that if u is a Dirichlet eigenfunction with eigenvalue λ then setting $r^2(x) = \sum_{i=1}^n x_i^2$ we have

$$\lambda = \frac{\int_{\partial\Omega} |\nabla u \cdot \hat{n}|^2 |\nabla r^2 \cdot \hat{n}|^2\, ds}{4\int_\Omega u^2\, dx}.$$

But since $\nabla u \cdot \hat{n} = 0$ due to the Neumann condition, we get that $\lambda = 0$ so from Lemma 2.1.10 $u \equiv 0$. □

Finally, we prove a rather unique result that gives us the location of zeroes of Laplacian eigenfunctions when $\Omega = [-\pi, \pi]$

Theorem 2.1.12 (Sturm Comparison Theorem [10]). *Let $\Omega = [a, b] \subseteq \mathbb{R}$ with $u + \lambda u = 0$ and $v + \mu v = 0$ where $\lambda > \mu$. Then u vanishes at least once between the consecutive zeroes of v.*

Proof. We compute the Wronskian, $W(u, v, x) = u'v - uv'$. Assume that a, b are consecutive zeroes of v such that $v(t) \neq 0 \in (a, b)$. Note that we may assume without losing generality that $v'(a) \geqslant 0$ and $v'(b) \leqslant 0$. Suppose towards a contradiction that u does not vanish on (a, b). Then $W(u, v, a) = u(a)v'(a) \geqslant 0$ and $W(u, v, b) = u(b)v'(b) \leqslant 0$. However, $W'(a, b, t) = uv'' - u''v = (\lambda - \mu)uv > 0$ so W is increasing. This is a contradiction since $W(b) \leqslant 0$. As a consequence, u had to vanish in (a, b). □

Example 2.1.2. Consider $\Omega = [-\pi, \pi]$. Then for any $u \in C^2$, $\Delta u = u''$. Furthermore, notice that $\{\sin(nx)\}$ and $\{\cos(nx)\}$ for $n \in \mathbb{N}$ are the Dirichlet and Neumann eigenfunctions of Δ respectively. In fact, from the uniqueness of second order linear ordinary differential equations, we have that these are the only eigenfunctions given the boundary conditions. As a consequence, from Theorem 2.1.9, we get that for any $f \in L^2([-\pi, \pi])$, $f = \sum_{n=0}^{\infty} a_n \cos(nx) + b_n \sin(nx) = \sum_{n \in \mathbb{Z}} c_n e^{inx}$, where equality holds in the L^2 sense. Additionally, from Corollary 2.1.4 we get that $\sin(nx)$ and $\cos(nx)$ are non-localized. Finally, with Theorem 2.1.12, we have that $\sin((n+1)x)$ vanishes at least as many times as the interior zeroes of $\sin((n+1)x)$. Observing that $\sin(n+1)x$ vanishes on the endpoints of the interval, we prove that $\sin(n+1)x$ has two more zeroes than $\sin(nx)$. With the fact that $\sin(x)$ vanishes thrice on $[-\pi, \pi]$, we conclude via induction that $\sin(nx)$ vanishes at least $2n+1$ times. A similar argument shows that $\cos(nx)$ vanishes at least $2n$ times. Thus, the three properties attributed to complex exponentials at the start of the section are really the analytic properties of the spectra of Laplacians.

As a concluding remark to this section, we pause here to reflect on why Sobolev Spaces are an appropriate setting to study elliptic equations like the Dirichlet problem. First, it is much easier to show that a function has a weak derivative than it is to show that it has a classical one. Second, proving the existence, orthogonality, and completeness of eigenfunctions is easier in the general setting of Hilbert Spaces than in the setting of continuously differentiable functions.

2.2 Spectral Graph Theory

In this section we introduce the essential tools of Spectral Graph Theory. After establishing the fundamental definitions on graphs, we introduce the normalized graph Laplacian $\mathcal{L}$ acting on $\mathbb{R}^n$ as the analogue of Δ on $L^2([-\pi, \pi])$. The bridge between Δ and $\mathcal{L}$ is the spectral theorem: $\mathcal{L}$, like Δ, is a symmetric operator on a Hilbert space of functions on graphs so it has an orthogonal basis of eigenvectors. Since the complex exponentials formed the Fourier basis for $L^2([-\pi, \pi])$, we interpret the eigenvectors of $\mathcal{L}$ as a Graph Fourier Basis for the graph G. In the second half of the section, we find that the Graph Fourier

Basis follows a weaker version of the Sturm Comparison Principle. Furthermore, we show that Corollary 2.1.4 is violated in the graph setting by providing examples of localized eigenfunctions. We end this chapter with a result connecting Corollary 2.1.4 and Theorem 2.1.11 in the graph setting: the occurrence of localized eigenfunctions is equivalent to finding a joint Dirichlet-Neumann eigenfunction on a graph. Most definitions and theorems stated in this section can be found in [6] and [3]. Propositions 2.2.1, 2.3.1, and 2.3.4 are novel. The idea for the proofs of Theorems 2.3.3 and 2.3.2 are suggested in [6, Chapter 8] but here we fill in the actual details.

Definition 2.2.1. Let $n \in \mathbb{N}$. A *finite undirected weighted graph* of n vertices is the ordered set $G = (V, E, w)$ where $V = [n] := \{1, \dots, n\}$ and $E \subseteq \binom{[n]}{2}$. The elements of V are the *vertices*, the elements of E are the *edges*, and $w : E \mapsto \mathbb{R}^+$ is the *weight function* which assigns a non-negative real number to every edge.

From now on, we shall assume that a graph is finite, undirected, and weighted unless otherwise specified. An unweighted graph G is a weighted graph where $w(e) = 1$ for every $e \in E$.

Definition 2.2.2. Let G be a graph and $A_G \in M_n(\mathbb{R})$ where $[A_G]_{ij} = w_{ij} \iff i, j \in E$ and 0 otherwise. Then A_G is the *adjacency matrix*. Let $T_G \in M_n$ be a diagonal matrix such that $[T_G]_{ii} = [A_G \mathbf{1}_n]_i$. Then T_G is the *degree* (or *weight*) matrix. The *degree* or *weight of a vertex* i is $d_i = [T_G]_{ii}$. We will drop the subscript G whenever it is clear which graph we are referring to.

Definition 2.2.3. Let G be a graph and let $f : V \mapsto \mathbb{R}$. Then f is a function on a graph.

Remark 2.2.1. Note that the space of functions on G is finite dimensional because V is finite. Consequently, any function $f : V \mapsto \mathbb{R}$ can be represented by an element in $\mathbb{R}^n$. For the remainder of this thesis, we will abuse notation by writing f as a function on V or as an element of $\mathbb{R}^n$ whenever convenient. This is justified because they are really the same thing: $f(x_i)$ is just the ith element in the vector $f \in \mathbb{R}^n$.

Now we attempt to mimic the theory in Section 2.1 by finding a meaningful operator that admits an orthogonal eigenbasis. What could this operator be? We derive it by going

back to the example from Section 1.3 from Chapter 1. Recall that we were given a graph Γ of the image and the problem was to detect the pixels that corresponded to the face. Let the set of face pixels be denoted F and the set of non-face pixels be B. We assume that all the pixels in F are close together and have similar pixel values. Recall that the edge weight between two pixels (or vertices) i and j was

$$w(i,j) = \exp\Big(\frac{-||x_i - x_j||_2^2}{2\sigma_x}\Big)\exp\Big(\frac{-||p_i - p_j||_2^2}{2\sigma_p}\Big). \tag{2.2.1}$$

As a consequence, we expect the edge weight between vertices in F to be large and that between F and B to be small. The problem of detecting face pixels could thus be reformulated as partitioning V into F and B such that the total edge weight between F and B is minimized. We convert this to a convex optimization problem.

Define the *total normalized edge weight* between any two partitions F and B of V as the cut:

$$\text{cut}(F,B) = \left(\frac{1}{\text{vol}(F)} + \frac{1}{\text{vol}(B)}\right)\sum_{i\in F, j\in B} w_{ij},$$

where $\text{vol}(F) = \sum_{i\in F} w_i$. Now consider a function $f : V \mapsto \mathbb{R}$ where

$$f(x) = \begin{cases} \frac{1}{\text{vol}(F)} \text{ if } x \in F, \\ -\frac{1}{\text{vol(B)}} \text{ if } x \in B. \end{cases}$$

Then

$$\text{cut}(F,B) = \Big(\frac{1}{\text{vol}(F)} + \frac{1}{\text{vol}(B)}\Big)\sum_{i\in F, j\in B} w_{ij} = \frac{\sum_{ij\in E} w_{ij}(f(i) - f(j))^2}{\sum_{i\in V} w_i (f(i))^2}.$$

Note that f takes values $\text{vol}(F)^{-1}$ or $\text{vol}(B)^{-1}$ but to minimize $\text{cut}(F,B)$ we may consider a "relaxation" of the above problem to all functions f on Γ such that $f \perp T\mathbf{1}$

$$\min_{F,B} \text{cut}(F,B) = \min_{f\perp T\mathbf{1}} \frac{\sum_{ij\in E} w_{ij}(f(i) - f(j))^2}{\sum_{i\in V} w_i (f(i))^2}.$$

The fraction on the right is a Generalized Rayleigh quotient for a symmetric matrix L since the numerator is a quadratic form and every quadratic form is associated with a symmetric matrix and the numerator. Furthermore, $\nabla\langle x, Lx\rangle = Lx$ so we can recover this matrix using the gradient of the numerator. This matrix is called the *Laplacian matrix*. The

actual calculation reveals that $L = T - A$. We can bring the Generalized Rayleigh quotient into standard form using a substitution:

$$\min_{f \perp T\mathbf{1}} \frac{\sum_{ij \in E} w_{ij}(f(i) - f(j))^2}{\sum_{i \in V} w_i (f(i))^2} = \min_{f \perp T\mathbf{1}} \frac{f^\top L f}{f^\top T f} = \min_{f \perp T\mathbf{1}} \frac{f^\top L f}{f^\top T f} = \min_{f \perp T^{1/2}\mathbf{1}} \frac{y^\top \mathcal{L} y}{y^\top y},$$

where $y = T^{1/2} f$ and $\mathcal{L} = T^{-1/2} L T^{-1/2}$. Incidentally, $T^{1/2}\mathbf{1}$ is actually an eigenvector of $\mathcal{L}$ so the above problem is an eigenvalue problem and the minimizer is an eigenvector.

Definition 2.2.4. Given a (finite, unweighted, undirected) graph $G = (V, E)$ and a function $u : V \mapsto \mathbb{C}$ the *Combinatorial Graph Laplacian* Δ_G is defined as:

$$\Delta_G u(x) = \sum_{y \sim x} (u(x) - u(y)).$$

Δ_G is a linear operator on the space of functions on V which, due to V's finitude, is isomorphic to $\mathbb{C}^{|V|}$ so it is accompanied by the Laplacian matrix L:

$$L_G = T_G - A_G.$$

Again, for notational clarity, we drop the subscript G whenever the graph is undisputed. The *Normalized Laplacian* [6], $\mathcal{L}$ is the matrix

$$\mathcal{L} = T^{-1/2} L T^{-1/2} = I - T^{-1/2} A T^{-1/2}.$$

The entries of $\mathcal{L}$ are

$$\mathcal{L}_{i,j} = \begin{cases} 1/d_i, & i = j, \\ -1/\sqrt{d_i d_j}, & i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\mathcal{L}$ slightly alters the action on a function u:

$$\mathcal{L}u(x) = \frac{1}{\sqrt{d_x}} \sum_{y \sim x} \left(\frac{u(x)}{\sqrt{d_x}} - \frac{u(y)}{\sqrt{d_y}} \right).$$

Since $\mathcal{L}$ is a symmetric matrix, the eigenvalues of $\mathcal{L}$ are characterized by the Rayleigh quotient and the respective minimax formulations (noting that the spectral decomposition of $\mathcal{L}$ is $\mathcal{L} = V \Lambda V^*$)

$$\lambda_0 = \min_{x \neq 0} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} = \frac{\langle x, \mathcal{L}x \rangle}{\langle T^{-1/2}x, T^{-1/2}x \rangle}, \tag{2.2.2}$$

$$\lambda_i = \min_{x \perp S_{i-1}} \frac{\langle x, \mathcal{L}x \rangle}{\langle T^{-1/2}x, T^{-1/2}x \rangle} \tag{2.2.3}$$

$$= \min_{x \neq 0} \max_{g \in S_{i-1}} \frac{\langle x, \mathcal{L}x \rangle}{\langle (x - g), (x - g) \rangle}. \tag{2.2.4}$$

Definition 2.2.5. Let G be a graph and $\mathcal{L}$ its Laplacian. Then $\mathcal{L} = V\Lambda V^*$ where V is an orthogonal matrix. The *Graph Fourier Basis* is the set of columns of V and the *Graph Fourier Transform* is the operator V^*. Since $L = T^{1/2}V\Lambda V^*T^{1/2}$, the *Combinatorial Fourier Basis* is the set of column vectors of $T^{1/2}V$. Note that the combinatorial Fourier Basis is NOT identical to the eigenbasis of the combinatorial Laplacian.

Remark 2.2.2. There are several reasons why calling L the "Laplacian" is appropriate. One of those reasons is because P_n the cyclic graph on n vertices can be viewed as the discretization of an interval on n points. In this case, the second-difference matrix used for numerically approximating second derivatives (i.e the Δ) is equal to the Laplacian on P_n. Hence, the Laplacian operator is a generalized "discrete" version of the Euclidean Laplacian.

Next, we prove a result on how many times a combinatorial Laplacian eigenfunction might oscillate.

Definition 2.2.6. Let $\mathcal{E}_f = \{(x_i, x_j) \in E \mid f(x_i)f(x_j) \leqslant 0\}$ be the *Nodal edges* of f and $\mathcal{V}_f = \{i \in V \mid f_i = 0 \text{ and } f_jf_k < 0 \text{ for some } j, k \sim i\}$ be the *Nodal vertices*. Then $\mathcal{N}_f = \mathcal{E}_f \cup \mathcal{V}_f$ is the *Nodal set* of f.

Note that if all edges in a graph are nodal, then the graph is bipartite.

Proposition 2.2.1. *Let f be a Combinatorial Laplacian λ-eigenfunction. Then $|\mathcal{E}_f| \geqslant G(\lambda)/2$ where $G(\lambda) = |\{x_i \in V \mid \sum_{x_j \sim x_i} w_{ij} < \lambda\}|$.*

Proof. The proof follows by reflecting on the eigenfunction formula and counting conservatively. Since f is an eigenfunction we have

$$\lambda f_i = \sum_{i\sim j} w_{ij}(f_i - f_j) = f_i \sum_{i\sim j} w_{ij} - \sum_{i\sim j} w_{ij}f_j.$$

Rearranging the above equation we have that

$$\sum_{i\sim j} w_{ij}f_j = \Big(\sum_{i\sim j} w_{ij} - \lambda\Big)f_i = A_{\lambda,i}f_i.$$

When $A_{\lambda,i} < 0$, $\sum_{i\sim j} w_{ij}f_j$ has a sign opposite to f_i. Not all f_j can have the same sign as f_i so there must be at least one $k \sim i$ such that $f_if_k < 0$ so $(i, j) \in \mathcal{E}_f$. We can do this

process for every vertex such that $A_{\lambda,i} < 0$ and collect the corresponding nodal edges. Note that we might have double counted some edges if $\exists\, i_1, i_2$ s.t. $i_1 \sim i_2$ and $A_{\lambda,i_1}, A_{\lambda,i_2} < 0$. So the number of distinct collected edges is at least half the number of collected edges (this is the most conservative case when every edge is double counted). But the number of counted edges is equal to the number of vertices for whom $A_{\lambda,i} < 0$, i.e $G(\lambda)$. Thus, $|\mathcal{E}_f| \geqslant G(\lambda)/2$. □

Remark 2.2.3. Since nodal edges connect positive vertices to negative ones, they describe how many times a function changes sign on the graph. Consequently Proposition 2.2.1 is analogous to Example 2.1.2 which establishes a lower bound on the number of times of Laplacian Eigenfunctions change sign on $[-\pi, \pi]$.

2.3 Localization Phenomena in Combinatorial Laplacian Eigenfunctions

Here we work with the combinatorial Laplacian, L.

Definition 2.3.1. Let $u : V \mapsto \mathbb{C}$ such that $Lu = \lambda u$ for some $\lambda \in \mathbb{R}$. Then u is *localized* if it vanishes on some neighbourhood of G. In other words, u is localized $\iff$ $\exists\, v \in V$ such that $u\,|_{N(v)} \equiv 0$

Example 2.3.1. The Dirichlet eigenfunctions on the graph approximations to the Sierpinski gasket exhibit localization, for example the 6-eigenfunction on the following graph:

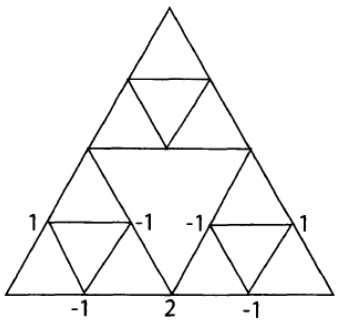


Figure 2.2: A Localizing 6-eigenfunction [29]. The vertices are the intersections of the segments and the numbers represent the function value at the vertices. No number means the value is zero.

Example 2.3.2. The *n-fan graph* is the graph F_n with $2n+1$ vertices $\{v_0, \ldots, v_{2n}\}$, and edges $\{v_0, v_i\}$ and $\{v_{2i}, v_{2i-1}\}$ for $1 \leqslant i \leqslant n$. F_1 is just the triangle with vertices v_0, v_1, and v_2, with eigenvectors $(1,1,1), (0,1,-1)$, and $(1,0,-1)$. It admits one 0-eigenfunction, one $(2n+1)$-eigenfunction, n 3-eigenfunctions and $n-1$ 1-eigenfunctions, adding up to $2n+1$ eigenfunctions, thus completing the spectrum. Remarkably, we have that almost all of the eigenvectors are localized on either one or two of the blades of the fan, so F_n is indeed a highly localizing graph.

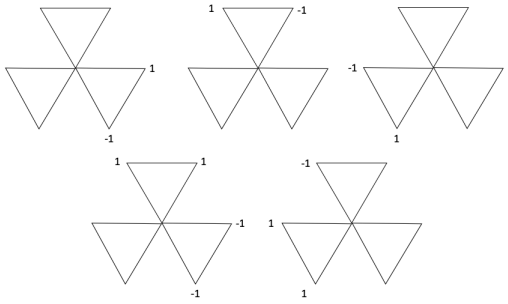


Figure 2.3: 3 and 1 eigenfunctions on the 3-fan

Example 2.3.3. The *n-star graph* is the graph S_n with $n+1$ vertices $\{v_0, \ldots, v_n\}$ and edges $\{v_0, v_i\}$ for $1 \leqslant i \leqslant n$. The spectrum of S_n can be deduced easily from the techniques used for F_n. Consider the contraction on $V(F_n)$ where v_{2i} and v_{2i-1} are contracted into one vertex. Then $S_n \equiv F_n/\sim$ where $\sim$ represents the contraction. The eigenvectors w_i are preserved in the contraction since each w_i takes the same value on each pair of contracted vertices. Consequently, we get $n-1$ 1-eigenfunctions for free. Furthermore, the $2n+1$ eigenfunction now becomes an $(n+1)$-eigenfunction. Combining these with the 0-eigenfunction completes the spectrum.

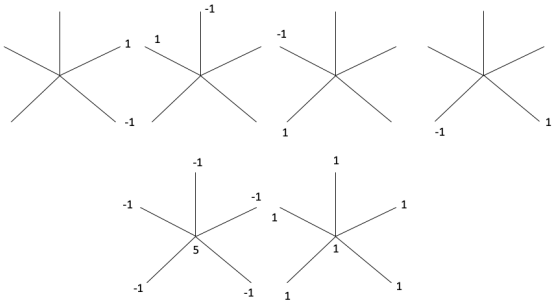


Figure 2.4: The spectrum of the 5-star graph

Remark 2.3.1. Note that although the eigenfunctions are localized, they are sometimes not orthogonal, such as the ones found in the 1-eigenspace.

Example 2.3.4. Suppose G_1 and G_2 are two graphs with eigenvectors u_1 and u_2 corresponding to eigenvalues λ_1 and λ_2 respectively such that u_1 vanishes at some vertex v_1 and u_2 vanishes at some vertex v_2. Let H be the graph formed by gluing G_1 and G_2 at the vertices v_1 and v_2. Then H admits a λ_1 and λ_2 eigenvector: one localized on G_1 and the other localized on G_2.

We can generalize the above observation to find a gluing decomposition for graphs with localization.

Definition 2.3.2. We say that a graph G is partially labelled if $\exists\, S \subseteq V(G)$ and a function $\sigma : S \mapsto \mathbb{N}$. Here σ is a *labelling* of the vertices.

Definition 2.3.3. Let G and H be partially labelled graphs with labellings σ_G and σ_H. Note that $G \sqcup H$ has a natural labeling σ. Let $\sim$ be the equivalence relation on $V(G \sqcup H)$ where $v \sim u \iff \sigma(v) = \sigma(u)$. Then the *gluing* of G and H is the graph $F \cong G \sqcup H/\sim$. In this case we use the notation denote $F = GH$.

Remark 2.3.2. This definition of gluing follows from [20], where the author notes that the operation is also associative and commutative. It is the multiplication operation in the gluing algebra of graphs.

Proposition 2.3.1. *Let F be a connected graph which admits a localized Laplacian λ-eigenvector u. Then $\exists\, G, H \subseteq F$ such that $F = GH$ where $\Delta_G u \mid_G = \lambda u \mid_G$.*

Proof. We partition F into a graph where u localizes and where it vanishes and show that F is the gluing of those two graphs at vertices where u vanishes but doesn't vanish in their neighbourhoods. To that end define $\mathcal{V} = \{v \in V(F) \mid u \mid_{N(v)} \equiv 0\}$. Setting $V(G) = \bigcup_{v\in\mathcal{V}} N(v)$, let $\partial G := \{v \in V(G) \mid \exists u \in V(G) \text{ and } u' \in V(F)\backslash V(G) \text{ s.t. } v \sim u, u'\}$. Finally, set G to be the subgraph induced by $\mathcal{V} \cup \partial G$ and H to be the subgraph induced by $\mathcal{V}^c$. Clearly, $F = GH$. Furthermore, if $v \in \mathcal{V}^c\backslash\partial G$ then v has neither a localizing

neighbourhood nor is it adjacent to a vertex with a localizing neighbourhood. So

$$\Delta_F u(v) = \sum_{x\sim v}(u(v)-u(x)) = \sum_{x\sim v, x\in\mathcal{V}^c}(u(v)-u(x)) = \Delta_G u(v).$$

Furthermore, if $v \in \partial G$ then

$$\begin{aligned}\Delta_F u(v) &= \sum_{x\sim v}(u(v)-u(x)) \\ &= \sum_{x\sim v, x\in\mathcal{V}}(u(v)-u(x)) + \sum_{x\sim v, x\in\mathcal{V}^c}(u(v)-u(x)) \\ &= \sum_{x\sim v, x\in\mathcal{V}^c}(u(v)-u(x)) \\ &= \Delta_G u(v).\end{aligned}$$

Thus, $\Delta_G u(v) = \lambda u(v)$ for every $v \in G$. □

Remark 2.3.3. First observe that ∂G is a vertex cut for F. Secondly, the above proposition can be interpreted in terms of the joint Dirichlet-Neumann spectrum of F because u is actually a joint Dirichlet-Neumann λ-eigenfunction on the subgraph induced by $\mathcal{V}^c\backslash\partial G$ with boundary ∂G. The terminology is borrowed from [6, Chapter 8].

2.3.1 Neumann and Dirichlet Spectra of Subgraphs

In this section we recap [6, Chapter 8]. Note that we now work with the normalized Laplacian but the techniques can be generalized to the combinatorial, or indeed any, Laplacian form on graphs.

Definition 2.3.4. Let G be a finite graph and $S \subseteq V(G)$. Then the *boundary vertices of* S, denoted δS, are the vertices of G which are not in S but adjacent to some vertex in S. Furthermore, the *boundary edges of* S, denoted ∂S are the edges which connect boundary vertices to S. Thus $\delta S = \{v \in V(S)^c \mid v \sim u \text{ for some } u \in V(S)\}$ and $\partial S = \{e \in E(G) \mid e = (x, y) \text{ such that } x \in V(S), y \in \delta S\}$. We also denote the union of the edges of S and its boundary edges as S^*. Of course, S may not have any boundary edges or any boundary vertices.

Definition 2.3.5. Let $S \subseteq G$. Then the *Neumann eigenfunction* f_1^N and *Neumann eigenvalue* λ_1^N of S are defined as follows:

$$\lambda_1^N = \min_{f \perp 1} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}, \qquad f_1^N = \operatorname*{argmin}_{f\perp 1} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}.$$

In general, given the subspace of the first $i-1$ Neumann eigenfunctions denoted N_{i-1}, the ith Neumann eigenfunction and eigenvalue are defined as follows:

$$\lambda_i^N = \min_{f \perp N_{i-1}} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}, \qquad f_i^N = \operatorname*{argmin}_{f\perp N_{i-1}} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}.$$

Definition 2.3.6. Let $S \subseteq G$. Then the *Dirichlet eigenfunction* f_1^D and *Dirichlet eigenvalue* λ_1^D of S are defined as follows:

$$\lambda_1^D = \min_{f|_{\delta S}\equiv 0} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}, \qquad f_1^D = \operatorname*{argmin}_{f|_{\delta S}\equiv 0} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}. \tag{2.3.1}$$

In general, given the subspace of the first $i-1$ Dirichlet eigenfunctions denoted D_{i-1}, the ith Dirichlet eigenfunction and eigenvalue are defined as follows:

$$\lambda_i^D = \min_{f \perp D_{i-1}} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}, \qquad f_i^D = \operatorname*{argmin}_{f\perp D_{i-1}} \frac{\sum_{(x,y)\in S*}(f(x)-f(y))^2}{\sum_{x\in V(S)}(f(x))^2 d_x}. \tag{2.3.2}$$

Remark 2.3.4. Note that these definitions directly replicate the minimax equations 2.2.2. However, the quotients found in equations 2.3.1 and 2.3.2 are not always the Rayleigh quotients of the restriction of the normalized Laplacian to S. For the Neumann case we find that the eigenvalues λ_i^N are the Rayleigh quotients of another symmetric operator given in [6, p. 125].

While we explicitly choose Dirichlet eigenfunctions to have vanishing boundary conditions, the definition for Neumann eigenfunctions does not a priori assert a vanishing normal derivative condition. The following two results prove that Neumann and Dirichlet eigenfunctions behave as we expect them to: that is, they mimic the properties of the Neumann and Dirichlet conditions on Euclidean domains with smooth boundaries.

Theorem 2.3.2. *Let $V(S) \subseteq V(G)$ and S the induced graph. The Neumann eigenfunction $f = f_1^N$ satisfies the following properties:*

1. *Fix $x \in V(S)$. Then*

$$\sum_{(x,y)\in S*} (f(x) - f(y)) = \lambda_N^1 d_x f(x). \tag{2.3.3}$$

2. *Fix $x \in \delta S$. Then*

$$\sum_{(x,y)\in \partial S} (f(x) - f(y)) = 0. \tag{2.3.4}$$

3. *Let $h : V(S) \cup \delta S \mapsto \mathbb{R}$. Then*

$$\sum_{x\in V(S)} h(x) L_{S\cup\delta S} f(x) = \sum_{(x,y)\in S*} (h(x) - h(y))(f(x) - f(y)). \tag{2.3.5}$$

Proof. We prove parts (1) and (2) variationally. For part (1), fix $x_0 \in V(S)$ and let

$$f_\epsilon(x) = \begin{cases} f(x_0) + \frac{\epsilon}{d_{x_0}} \text{ if } x = x_0, \\ f(x) - \frac{\epsilon}{vol(S) - d_{x_0}}. \end{cases}$$

First observe that $f_\epsilon \mid_S \perp T1_S$ so the minimization problem is well-defined on f_ϵ. Now we compute the quotient for f_ϵ:

$$\begin{aligned} R(\epsilon) &= \frac{\sum_{(x,y)\in S*} (f_\epsilon(x) - f_\epsilon(y))^2}{\sum_{x\in V(S)} (f_\epsilon(x))^2 d_x} \\ &= \frac{\sum_{(x,y)\in S*, x\neq x_0} (f(x) - f(y))^2 + \sum_{(x_0,y)\in S*} (f(x) + \frac{\epsilon}{d_{x_0}} - f(y) + \frac{\epsilon}{\mathrm{vol}(S) - d_{x_0}}))^2}{\sum_{x\neq x_0} (f(x) - \frac{\epsilon}{\mathrm{vol}(S)-d_{x_0}})^2 d_{x_0} + (f(x_0) + \frac{\epsilon}{d_{x_0}})^2 d_{x_0}} \\ &= \frac{\sum_{(x,y)\in S*} (f(x) - f(y))^2 + \frac{2\epsilon \mathrm{vol}(S)}{d_{x_0}(\mathrm{vol}(S)-d_{x_0})} \sum_{(x_0,y)\in S*} (f(x_0) - f(y)) + O(\epsilon^2)}{\sum_{x\in V(S)} (f(x))^2 d_x + \frac{2\epsilon f(x_0) d_{x_0} \mathrm{vol}(S)}{d_{x_0}(vol(S)-d_{x_0})} + O(\epsilon^2)}. \end{aligned}$$

The second equality follows after simplifying the algebra and noting that $\sum_{x\in V(S)} f(x) d_x = 0$. We know that when $\epsilon = 0$, $f_\epsilon = f$, which also minimizes $R(\epsilon)$. Thus, $R'(0) = 0$ so computing the derivative via the quotient rule and setting the numerator at $\epsilon = 0$ to zero, we get that

$$\begin{aligned} &\Big(\frac{2\mathrm{vol}(S)}{d_{x_0}(\mathrm{vol}(S) - d_{x_0})} \sum_{(x_0,y)\in S*} (f(x_0) - f(y))\Big) \sum_{x\in V(S)} (f(x))^2 d_x \\ &\quad - \Big(\frac{2 f(x_0) d_{x_0} \mathrm{vol}(S)}{d_{x_0}(\mathrm{vol}(S) - d_{x_0})}\Big) \sum_{(x,y)\in S*} (f(x) - f(y))^2 = 0. \end{aligned}$$

Rearranging the equation, dividing through by $\sum_{x\in V(S)} (f(x))^2 d_x$ and noting that

$$\frac{\sum_{(x,y)\in S^*} (f(x)-f(y))^2}{\sum_{x\in V(S)} (f(x))^2 d_x} = \lambda_1^N$$

gives us part (1). For part (2), we adopt a similar strategy but the variation is simpler. Fix $x_0 \in \delta S$ and set

$$f_\epsilon(x) = \begin{cases} f(x) + \epsilon \text{ if } x = x_0, \\ f(x) \text{ otherwise.} \end{cases}$$

Now compute the Neumann quotient and observe that we can separate the sum in the numerator over edges that connect to x_0 and those that don't. By definition, the edges that connect with x_0 are contained in ∂S so

$$\begin{aligned} R(\epsilon) &= \frac{\sum_{(x,y)\in S^*} (f_\epsilon(x) - f_\epsilon(y))^2}{\sum_{x\in V(S)} (f_\epsilon(x))^2 d_x} \\ &= \frac{\sum_{(x,y)\in S^*} (f(x)-f(y))^2 + 2\epsilon \sum_{(x_0,y)\in \partial S} (f(x_0) - f(y)) + O(\epsilon^2)}{\sum_{x\in V(S)} (f(x))^2 d_x}. \end{aligned}$$

Once again, taking the derivative with respect to ϵ and setting it 0 at $\epsilon = 0$ yields (2). Lastly, (3) is a consequence of (2) because

$$\begin{aligned} \sum_{x\in V(S)} h(x) L_{S\cup \delta S} f(x) &= \sum_{x\in V(S)} h(x) \sum_{(x,y)\in S^*} (f(x) - f(y)) \\ &= \sum_{x\in V(S)} h(x) \sum_{(x,y)\in S^*} (f(x) - f(y)) + \sum_{y\in \delta S} h(y) \sum_{\substack{x\sim y \\ x\in V(S)}} (f(y) - f(x)). \end{aligned}$$

If we observe the contribution of each edge (x, y) to the above sum we observe that the vertices get counted exactly twice, with the sum over the edge being

$$h(x)(f(x) - f(y)) + h(y)(f(x) - f(y)) = (h(x) - h(y))(f(x) - f(y)).$$

This completes the argument. □

An analogous result holds for Dirichlet eigenfunctions.

Theorem 2.3.3. *Let $V(S) \subseteq V(G)$ and S the induced graph. The Dirichlet eigenfunction $f = f_1^D$ satisfies the following properties:*

1. *Fix $x \in V(S)$. Then*
$$\sum_{(x,y)\in S*} (f(x) - f(y)) = \lambda_N^1 d_x f(x). \tag{2.3.6}$$
2. *Let $h : V(S) \cup \delta S \mapsto \mathbb{R}$ and $h(x) = 0$ for every $x \in \delta S$. Then*
$$\sum_{x\in V(S)} h(x) L_{S\cup\delta S} f(x) = \sum_{(x,y)\in S*} (h(x) - h(y))(f(x) - f(y)). \tag{2.3.7}$$

Proof. We get to use simpler variations as the constraint on the minimization problem is different. For part (1), we fix $x_0 \in S$ and use the following variation:
$$f_\epsilon(x) = \begin{cases} f(x) + \epsilon \text{ if } x = x_0, \\ f(x) \text{ otherwise.} \end{cases} \tag{2.3.8}$$
f_ϵ satisfies the vanishing boundary condition so we can compute the given Dirichlet quotient and then apply the same method as the proof for part (1) of Theorem 2.3.2. For part (2), notice that
$$\begin{aligned} \sum_{x\in V(S)} h(x) L_S f(x) - \sum_{(x,y)\in S*} (h(x) - h(y))(f(x) - f(y)) \\ = \sum_{\substack{(x,y)\in \partial S \\ y\in\delta S}} h(y)(f(x) - f(y)) = 0. \end{aligned}$$
□

Remark 2.3.5. The differences between Theorems 2.3.2 and 2.3.3 are subtle but important. The first important difference is that while for the Neumann result the bilinear form in part (3) is true for *any* function h while the one in part (2) of the Dirichlet theorem is true only for functions h vanishing on δS. Secondly, the proof for the Neumann condition involves perturbing f on the boundary; as a consequence, we cannot use the same technique for the Dirichlet eigenfunctions, which are fixed at 0 on δS. Lastly, both Dirichlet and Neumann functions satisfy the eigenvalue equation for $\mathcal{L}_{S\cup\delta S}$ on S. This does not mean that they are Laplacian eigenfunctions on $S \cup \delta S$.

Although the function f in Equations 2.3.1 and 2.3.2 does not satisfy an eigenvalue equation, the function $g = T^{1/2} f$ on S does. Additionally, the vanishing normal derivative

condition 2.3.4 changes to

$$\frac{1}{\sqrt{d'_x}} \sum_{\substack{(x,y)\in\partial S \\ y\in S}} \frac{g(x)}{\sqrt{d'_x}} - \frac{g(y)}{\sqrt{d_y}} = 0. \tag{2.3.9}$$

As a consequence g is an true eigenfunction of the normalized Laplacian with boundary conditions. Here we suggest an alternative formulation of these boundary problems in the language of operators. For the following definitions, assume $S \subseteq G$

Definition 2.3.7. The *Dirichlet Operator* (resp. *Normalized Dirichlet Operator*), D_S (resp. $\mathcal{D}_S$) is the restriction of L_G (resp $\mathcal{L}_G$ to the rows and columns of S).

Definition 2.3.8. The *boundary matrix* B is a $|\delta S| \times |S|$ matrix where B_{ij} is 1 if the ith boundary vertex is connected to the jth vertex in S.

Definition 2.3.9. The *boundary degree matrix* is the diagonal matrix δT_S where the ith diagonal entry is the degree of the ith boundary point into S.

Definition 2.3.10. The *Neumann Operator* (resp. *Normalized Neumann Operator*), N_S (resp. $\mathcal{N}_S$) are the following operators on the space of functions $f : S \mapsto \mathbb{R}$:

$$N_S = D_S - B^\top (\delta T_S)^{-1} B,$$
$$\mathcal{N}_S = T_S^{-1/2} N T_S^{-1/2}.$$

Remark 2.3.6. If f satisfies the Dirichlet condition on the boundary of S then taking the Laplacian on any vertex of S we get that

$$\Delta_G f(x) = d_x f(x) - \sum_{y\in S\cup\delta S} f(y) = d_x f(x) - \sum_{y\in S} f(y). \tag{2.3.10}$$

As a consequence, we can represent the action of the Laplacian on S on a function with Dirichlet conditions by a matrix equal to the rows and columns of L_G indexed by S. The Neumann situation is a bit more complicated. Suppose f satisfies the Neumann condition on the boundary. Then equation 2.3.10 still holds but the boundary terms are not zero. Instead, if $y \in \delta S$ then from the Neumann condition we have that $f(y) = (1/d'_y)\sum_{z\sim y} f(z)$. We can substitute this back into equation 2.3.10 and get that

$$\Delta_G f(x) = d_x f(x) - \sum_{y\in S\cup\delta S} f(y) = \underbrace{d_x f(x) - \sum_{\substack{y\in S \\ y\sim x}} f(y)}_{\text{Represented by } D} - \underbrace{\sum_{\substack{y\in\delta S \\ y\sim x}} \frac{1}{d'_y} \sum_{\substack{z\sim y \\ z\in S}} f(z)}_{\text{Represented by } B^\top(\delta T_S)^{-1}B} .$$

Proposition 2.3.4. *Let $S \subseteq V(G)$, δS the induced boundary, ∂S the boundary edges, and $S^* = \partial S \cup E(S)$. Furthermore, suppose u is a joint Dirichlet-Neumann eigenfunction on S. Then u can be extended to G such that u is localized on S.*

Proof. Letting $u = 0$ on $(V(S) \cup \delta S)^c$, we see that when $x \in (V(S) \cup \delta S)^c$,

$$\Delta_G u(x) = \sum_{x \sim y} (u(x) - u(y)) = 0.$$

because y can only be in δS (if it was in S then x would be in either δS or $V(S)$).

If $x \in \delta S$ then

$$\begin{aligned}
\Delta_G u(x) &= \sum_{x \sim y} (u(x) - u(y)) \\
&= \underbrace{\sum_{x \sim y, y \in V(S)} (u(x) - u(y))}_{=0 \text{ because of the Neumann condition}} + \underbrace{\sum_{x \sim y, y \in (V(S) \cup \delta S)^c} (u(x) - u(y))}_{=0 \text{ due to Dirichlet + Extension conditions}} \\
&+ \underbrace{\sum_{x \sim y, y \in \delta S} (u(x) - u(y))}_{=0 \text{ because of the Dirichlet condition}} \\
&= 0.
\end{aligned}$$

Finally, if $x \in V(S)$ then $\Delta_G u(x) = \lambda u(x)$ because u is a Dirichlet (or Neumann) eigenfunction. □

To end this chapter, we propose two open problems motivated by Remark 2.3.3, Proposition 2.3.1, and Proposition and 2.3.4:

Question 2.3.5. *Suppose G admits a localized eigenvector f. Is the cut induced by f minimal?*

Question 2.3.6. *Which graphs G admit localized eigenvectors? Equivalently which graphs G admit a subgraph S such that S has a joint Dirichlet-Neumann eigenvector?*

One approach to answering Question 2.3.6 is to notice that the problem is equivalent to asking when the kernel of $B^\top(\delta T_S)^{-1}B$ contains a Dirichlet eigenvector. The approach would be to deduce structural properties on S from the fact that $B^\top(\delta T_S)^{-1}B$ contains an eigenvector of D_S. Using this strategy, we can figure out some necessary conditions but the ones found by the author thus far are not sufficient.

Chapter 3

Diffusion Maps

In this chapter we discuss Diffusion Maps pioneered in [7, 8, 4] as a multiresolution framework for graphs. In [8], Coifman et al. proposed Diffusion Maps as a construction that generalizes the kernel methods found in machine learning. In 2012, Allard, Chen, and Maggioni [1] proposed a geometric multi-resolution analysis on graphs using tree decompositions and diffusion maps. These ideas merit an undergraduate thesis of their own; here we discuss Laplacian Eigenmaps pioneered by Belkin and Niyogi in 2003 as they were essential precursors to the ideas in [8]. As a preliminary remark we note that unlike the Fourier basis, Diffusion Maps are not typically used for decomposing signals. Instead, they resemble multiresolution analyses qualitatively because their properties enable detecting finer and coarser structures within graphs. These properties make them ideal for studying low-dimensional structures in high-dimensional data. This chapter is more experimental than the previous one. In the first section, we set up Belkin and Niyogi's construction and in the second section we show three examples to Graph Drawing, Manifold Identification, and Image resolution. In the third section, we provide a novel construction called *Neumann maps* by applying the process in Section 3.1 to the Neumann operator in 2.3.10. We end the thesis with some examples of Neumann maps which suggest that these are indeed worth studying.

3.1 Diffusion maps on graphs

Let G be a weighted, undirected, and finite graph. Consider a random walk on G where

$$P(X(t) = j \mid X(t-1) = i) = \frac{w_{ij}}{d_i}.$$

In other words, the property of jumping to vertex j from vertex i equals the relative weight of the ijth edge with respect to the total weight on the vertex i. The Law of total probability gives the probability of being at vertex j at time t:

$$P(X(t) = j) = \sum_i P(X(t) = j \mid X(t-1) = i)P(X(t-1) = i) = M[j,:]X(t-1),$$

where $M[j,:] = [P(X(t) = j \mid X(t-1) = 1) \dots P(X(t) = j \mid X(t-1) = n)]$. Let's make a matrix of these suggestively labelled row vectors, called M, where the jth row of M is $M[j,:]$. This is termed the probability transition matrix. Note that M^t gives us the probability transition at time t. Furthermore, observe that

$$M = AT^{-1},$$

where T is the degree matrix, W is the weight matrix. M isn't a symmetric matrix, but it is similar to one:

$$S = T^{-1/2}MT^{1/2} = T^{-1/2}WT^{-1/2} = I - \mathcal{L}. \tag{3.1.1}$$

We can write M in terms of the spectral decomposition of $I - \mathcal{L} = V\Lambda V^{-1}$ where $\Lambda = I - \Omega$ where Ω is the diagonal matrix of Normalized Laplacian eigenvalues and V is the Fourier Basis matrix.

$$M = T^{1/2}ST^{-1/2} = T^{1/2}V\Lambda V^{-1}T^{-1/2} = (T^{1/2}V)\Lambda(T^{1/2}V)^{-1} = \Phi\Lambda\Phi^{-1} = \Phi\Lambda\Psi^\top.$$

Here $\Phi = [\varphi_1, \dots, \varphi_n]$ and $\Psi = [\psi_1, \dots, \psi_n]$ so $\Psi^\top = [\psi_1^\top, \dots, \psi_n^\top]^\top$.We may further expand M in terms of φ_i and ψ_i in the following way:

$$M = \Phi\Lambda\Psi^\top = [\varphi_1, \dots, \varphi_n]\Lambda[\psi_1^\top, \dots, \psi_n^\top]^\top = \sum_{i=1}^{n} \lambda_i \varphi_i \psi_i^\top.$$

Since we let $M_{ij} = w_{ij}/dj$, each column of M is the probability vector of walking from the jth vector to anywhere in the graph in 1 step. Similarly, jth column of M^t denotes the

probability of walking from j at time t. So we may write the jth column of M^t as follows:

$$M^t[:, j] = \sum_{i=1}^{n} \lambda_i^t \varphi_i \psi_i^\top(j).$$

Note that φ_i form a basis for $\mathbb{R}^n$ so we can write the coordinate vector for $M^t[:, j]$ in that basis:

$$\xi_t(j) = \begin{bmatrix} \lambda_1^t \psi_1^\top(j) \\ \lambda_2^t \psi_2^\top(j) \\ \vdots \\ \lambda_n^t \psi_n^\top(j) \end{bmatrix}.$$

Finally, observe that the first component is redundant because the first eigenvalue is always 0. We have thus found an embedding of every vertex in $(n-1)$-dimensional space. This is called a **diffusion map**:

Definition 3.1.1 (Diffusion Maps). Let G be a weighted undirected graph. The *Diffusion Map* of G at time t is the function $\xi_t : V(G) \mapsto \mathbb{R}^{n-1}$ where

$$\xi_t(j) = \begin{bmatrix} \lambda_2^t \psi_2^\top(j) \\ \vdots \\ \lambda_n^t \psi_n^\top(j) \end{bmatrix}.$$

Definition 3.1.2 (d-dimensional Diffusion Map). Given a diffusion map ξ_t, the d-dimensional diffusion map is given by its projection on the subspace generated by the first d basis vectors:

$$\xi_t^{(d)}(j) = \begin{bmatrix} \lambda_2^t \psi_2^\top(j) \\ \vdots \\ \lambda_n^t \psi_d^\top(j) \end{bmatrix}.$$

Remark 3.1.1. The diffusion map is a continuous map with respect to the time parameter t. Furthermore, the Euclidean distance between two diffusion map points gives us the weighted difference between the distribution when starting from vertex i_2 and the distribution when starting from vertex i_2.

Theorem 3.1.1. $\sum_{j=1}^{n} (P(X(t) = j \mid X(0) = i_1) - P(X(t) = j \mid X(0) = i_2))^2 \frac{1}{d_j} = ||\xi_t(i_1) - \xi_t(i_2)||$.

Proof. We express the left hand side in matrix form and then compute:

$$\begin{aligned}
&\sum_{j=1}^{n}(P(X(t)=j \mid X(0)=i_1)-P(X(t)=j \mid X(0)=i_2))^2\frac{1}{d_j}\\
&=(M^t\chi_{i_1}-M^t\chi_{i_2})^\top T^{-1}(M^t\chi_{i_1}-M^t\chi_{i_2})\\
&=(M^t(\chi_{i_1}-\chi_{i_2}))^\top T^{-1}(M^t(\chi_{i_1}-\chi_{i_2}))\\
&=(\chi_{i_1}-\chi_{i_2})^\top (M^t)^\top T^{-1}M^t(\chi_{i_1}-\chi_{i_2})\\
&=(\chi_{i_1}-\chi_{i_2})^\top (\Phi\Lambda^t\Psi^\top)^\top T^{-1}\Phi\Lambda^t\Psi^\top(\chi_{i_1}-\chi_{i_2})\\
&=(\chi_{i_1}-\chi_{i_2})^\top \Psi\Lambda^t\Phi^\top T^{-1}\Phi\Lambda^t\Psi^\top(\chi_{i_1}-\chi_{i_2})\\
&=(\chi_{i_1}-\chi_{i_2})^\top \Psi\Lambda^t\underbrace{V^\top T^{1/2}T^{-1}T^{1/2}V^\top}_{I}\Lambda^t\Psi^\top(\chi_{i_1}-\chi_{i_2})\\
&=(\Lambda^t\Psi^\top(\chi_{i_1}-\chi_{i_2})^\top)^\top\Lambda^t\Psi^\top(\chi_{i_1}-\chi_{i_2})\\
&=||\xi_t(i_1)-\xi_t(i_2)||^2.
\end{aligned}$$

□

3.2 Applications of Diffusion Maps

By associating vertex j with the point $\xi_t^{(d)}(j)$ in $\mathbb{R}^d$ a diffusion map embeds the original graph G in a d-dimensional space. We term this a *diffusion embedding*. Furthermore, Theorem 3.1.1 demonstrates that the Euclidean distance between two points in the embedding is equal to their proximity in the random walk. This has major implications on the shape of the Diffusion Embedding because two vertices (or points) are placed close to each other if their edge-wise weights are large. Similarly, we may form a graph from any finite set of points using a known notion of pairwise affinity. For instance, we can associate a graph to an image where the affinity between two pixels is given by Equation 2.2.1. The Diffusion Map will then "draw" that set as a Diffusion Embedding. As such, a Diffusion Map is a local to global algorithm: it takes local information and gives a global drawing. This has several applications and here we highlight three in particular.

3.2.1 Graph Drawing

A diffusion embedding is a drawing of a graph in $\mathbb{R}^d$. How intuitive or useful is this drawing? The following result shows that the 2-dimensional diffusion embedding of a circulant graph

is the circular embedding in the plane, which is how we would intuitively draw a circulant graph (a graph whose adjacency matrix is circulant).

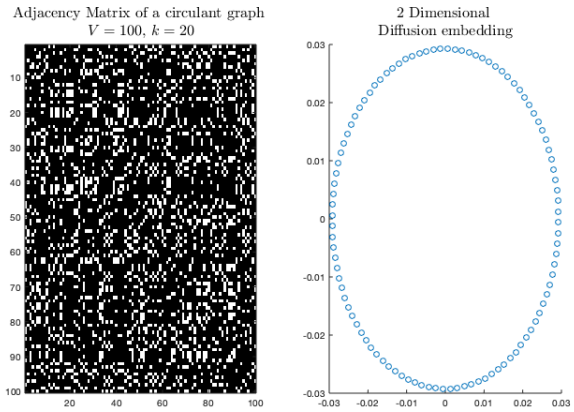


Figure 3.1: Left: The adjacency matrix of a 20-regular circulant graph on 100 vertices. Right: $\{\xi_2^1(j)\}_{1\leqslant j\leqslant 100}$. Note that we purposely represent the adjacency matrix in jumbled form (the rows are placed randomly instead of in the order where the matrix is circulant) to emphasize that the diffusion embedding

Proposition 3.2.1. *Let G be a circulant graph. Then the 2-dimensional diffusion maps $\{\xi_t^{(2)}(i)\}$ lie on a circle in $\mathbb{R}^2$.*

Proof. Since G is circulant, it is k-regular and $\mathcal{L}$ is circulant and symmetric. Consequently, $T^{1/2} = \sqrt{k}I$ and the eigenvectors are $y_m = \sin(2\pi mj/n)$ and $x_m = \cos(2\pi mj/n)$ where $0 \leqslant m \leqslant \lfloor n/2 \rfloor$ with x_m and y_m corresponding to the same eigenvalue. Then the 2nd and 3rd eigenvectors correspond to $m = 1$ with eigenvalue λ:

$$\psi_2^\top = \sqrt{k}[\cos(2\pi 0/n), \ldots, \cos(2\pi(n-1)/n)],$$

$$\psi_3^\top = \sqrt{k}[\sin(2\pi 0/n), \ldots, \sin(2\pi(n-1)/n)].$$

Consequently, $\xi_t^{(2)}(j) = \sqrt{k}[\lambda^t \cos(2\pi j/n)\, \lambda^t \sin(2\pi j/n)]^\top$ so the diffusion maps lie on a circle of radius $\sqrt{k}\lambda^t$ centered at the origin. □

3.2.2 Identifying Manifolds

A set of points (also referred to as a pointcloud) $X = \{x_i\}_{1\leqslant i\leqslant N} \subset \mathbb{R}^n$ can be associated with a graph G on N vertices where each vertex represents x_i and $w_{ij} = \exp\frac{-||x_i - x_j||^2}{\sigma}$ for some positive σ. Then $\{\xi_d^t(j)\}_{1\leqslant j\leqslant N}$ is another pointcloud of size N but is located in $\mathbb{R}^d$.

This is especially useful when d is much smaller than n because the diffusion embedding can extract features out of pointclouds which might be situated in high-dimensional Euclidean space but require a small number of parameters to describe (eg.: Points sampled from a circle in $\mathbb{R}^{100}$ have 100 components but are all described by one parameter: angle with the center). As a trivial example, if we sample points from a torus, the diffusion embedding is a torus.

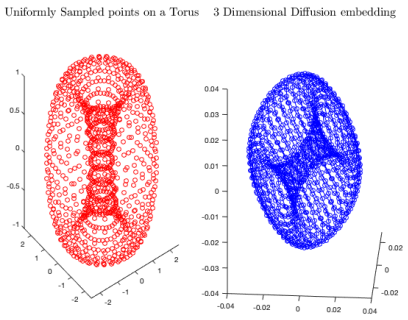


Figure 3.2: Left: A pointcloud sampled uniformly from a torus. Right: The 3-dimensional diffusion embedding of the graph associated with the pointcloud is a torus

3.2.3 A Simple Multiresolution Framework

The dimension d and scale t of $\xi^t_{(d)}$ provide a natural time-space resolution of the graph. This can be implemented to detect submanifolds within an image.

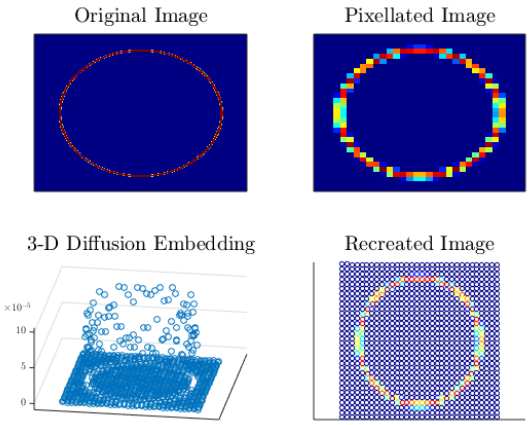


Figure 3.3: We compute the diffusion embedding of Γ_I, the graph associated with the image of a circle on a plain background. We compute $\{\xi^t_d\} for$d=3, t=1.

At first glance, Figure 3.3 seems to be a redundant redrawing of the original image. But note that the diffusion map did not require co-ordinates at all: it was able to find them on its own! More pertinently, the first two components of $\xi_1^{(3)}$ form a lattice, which is exactly the geometry we use to embed the graph of an image. Furthermore, the third component was used in "resolving" the submanifold within the graph. Thus, in the first two dimensions the diffusion map resolved the ambient manifold and used the third dimension to resolve the submanifold.

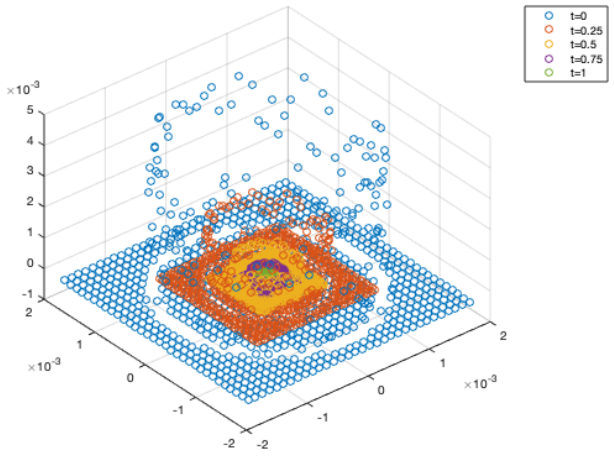


Figure 3.4: Effect of scale on the Diffusion Embedding

From Figure 3.4 we see that scales are rudimentary in this case as higher scales do not necessarily yield more novel representations. In this way, the Diffusion map is similar to the Littlewood-Paley form of the Fourier series in the sense that it has one important parameter which it uses to globally resolve the image. However, we know that sometimes localized frameworks (such as the Haar Wavelet) can prove particularly useful in detecting features in specific regions of the image. In the next section, we propose an idea for such an adaptive Diffusion framework.

3.3 Neumann Maps

To construct a localized framework, we understand a subgraph as a subdomain within a graph. We may treat the subgraph S as a graph in itself but doing so might erase important

information from the ambient graph via the boundary δS. The goal is to then find a suitable construction of a Diffusion Map which takes boundary information into account. In the previous section we saw that all we need for a diffusion map is a random walk which can be "symmetrized" via something like equation 3.1.1. Conversely, we can take a symmetric operator and then reverse the symmetrization process to form a right or left stochastic matrix. To do so, we turn to the Normalized Neumann operator $\mathcal{N}$. We suppose that our random walk matrix R is similar to the identity minus Normalized Neumann operator, that is $I - \mathcal{N} \sim R$. From equation 3.1.1 we have that $M = T^{1/2}(I - \mathcal{L})T^{-1/2}$. Analogously, set $R = T_S^{-1/2}(I - \mathcal{N})T_S^{1/2}$. The following proposition proves that $R^\top$ is row stochastic so we were justified in the choice of operator.

Proposition 3.3.1. *Let $R = T_S^{1/2}(I - \mathcal{N})T_S^{-1/2}$. Then $R^\top \mathbf{1} = \mathbf{1}$ and the entries of R are all non-negative.*

Proof. This follows because the Neumann matrix admits $T^{1/2}\mathbf{1}$ as a zero-eigenvector:

$$R^\top \mathbf{1} = T_S^{-1/2}(I - \mathcal{N})T_S^{1/2} = I\mathbf{1} - T_S^{-1/2}\mathcal{N}T_S^{1/2}\mathbf{1} = I\mathbf{1} = I\mathbf{1} = \mathbf{1}.$$

To see the non-negativity of the entries, we expand N in terms of the Dirichlet and Boundary operators:

$$\begin{aligned} R^\top &= T_S^{-1/2}(I - \mathcal{N})T_S^{1/2} \\ &= I - T_S^{-1/2}\mathcal{N}T_S^{1/2} \\ &= I - T_S^{-1/2}(\mathcal{L}_\int - T_S^{-1/2}B^\top(\delta T_S)^{-1}BT_S^{-1/2})T_S^{1/2} \\ &= I - T_S^{-1/2}\mathcal{L}_S T_S^{1/2} + T_S^{-1}B^\top(\delta T_S)^{-1}B \\ &= M_S^\top + T_S^{-1}B^\top(\delta T_S)^{-1}B. \end{aligned}$$

Here $M_S^\top$ represents the rows and columns of $M^\top$ indexed by S. Since M is a left stochastic matrix its entries are non-negative. Furthermore, the entries of $T_S^{-1}B^\top(\delta T_S)^{-1}B$ are also non-negative so R is indeed column stochastic. □

Remark 3.3.1. The matrix R is the transition matrix of the reflecting random walk mentioned in [6]. The reflecting walk proceeds as follows: if $u \in S$ then the probability of

walking to any neighbour of u is $(1/d_u)$. If $x \in \delta S$ such that $x \sim u$ then the walk proceeds to a neighbour $y \sim x$ where $y \in S$ with probability $(1/d_u d'_x)$. As a consequence, there is a non-zero probability of remaining stationary because there is a chance that the walk could reflect off the boundary back onto the original vertex.

We mimic the construction in Section 3.1 to define Neumann diffusions.

Definition 3.3.1. Let R be the reflecting walk matrix where $R = (T_S^{1/2}W)\Sigma(T_S^{1/2}W)^{-1} = C\Sigma B^\top$. Then the Neumann diffusion map of vertex $i \in S$ is the following point in $\mathbb{R}^{|S|-1}$:

$$g_t(i) = \begin{bmatrix} s_2 b_1^\top(i) \\ \vdots \\ s_n b_n^\top(i) \end{bmatrix}.$$

Similarly, the d-dimensional Neumann diffusion map is the projection of $g_t(i)$ to the standard d-dimensional subspace of $\mathbb{R}^{|S|-1}$:

$$g_t^{(d)}(i) = \begin{bmatrix} s_2 b_1^\top(i) \\ \vdots \\ s_d b_d^\top(i) \end{bmatrix}.$$

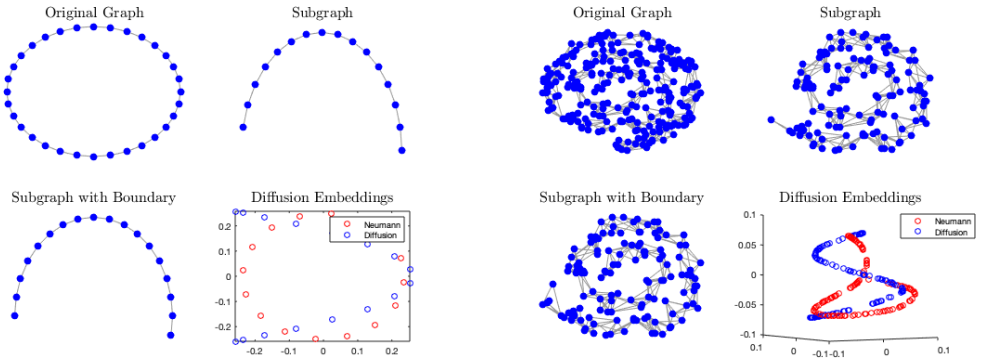
Here s_i are the eigenvalues of R contained in the diagonal matrix Σ and W are the Neumann eigenvectors. Note that $s_i = 1 - \rho_i$ where ρ_i is the ith Neumann eigenvalue.

Let $Q(X(t) = i \mid X(0) = j)$ be the probability of walking to vertex i through a reflecting random walk after starting at vertex j. Then the analogue of Theorem 3.1.1 holds:

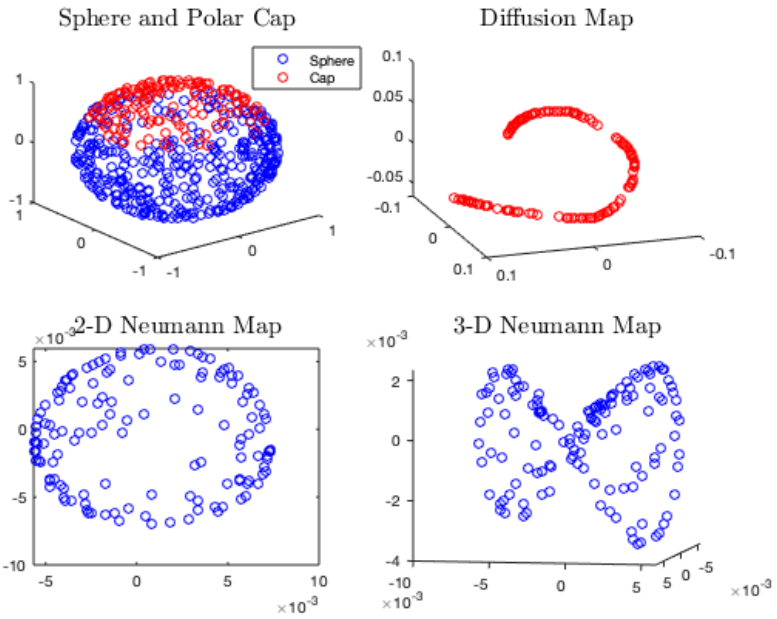
Theorem 3.3.2. $\sum_{j=1}^{|S|}(Q(X(t) = j \mid X(0) = i_1) - Q(X(t) = j \mid X(0) = i_2))^2 \frac{1}{d_j} = ||g_t(i_1) - g_t(i_2)||$.

Figures 3.5a and 3.5b provide experimental evidence for why Neumann maps might be more refined tools for resolving submanifolds. The most compelling example is given by Figure 3.5b where a 3 dimensional Diffusion Embedding of a polar cap as a subset of a sphere is a helix; whereas the Neumann map intelligently embeds it as a surface instead. We would think that the Neumann embedding is more intuitive since the sub-point cloud comes from a surface and not a curve. In fact the 2-D Neumann map resolves the cap

as a disk (which is resolved by two parameters) while the Diffusion map requires at least 3 dimensions to resolve the two-parameter dependence of the pointcloud. This is because the reflecting random walk "informs" the embedding of the ambient manifold while the standard random walk does not contain that information.



(a) Neumann and Diffusion embeddings of a as a subgraph of a circle and a spiral with two twists as a subgraph of a spiral with three



(b) Neumann and Diffusion embeddings of a polar cap as a subset of a sphere

Figure 3.5: Comparing Neumann and Diffusion Embeddings for subgraphs three graphs: Cycle, Spiral, and Sphere

Chapter 4

Conclusion

We take a moment to briefly reflect on what has conspired in the previous 7^2 pages. In Chapter 1 we demonstrated through Haar's example that MRAs form the natural setting to encode the three properties of completeness, resolution, and localization important in image processing. In Chapter 2 we revisited these properties in the Fourier basis on $[-\pi, \pi]$ and demonstrated that they emerge as consequences of the spectral and regularity theory of Laplacian eigenfunctions. This motivated the use of the eigenbasis of the graph Laplacian as a Graph Fourier basis (GFB). Proposition 2.2.1 showed that the GFB in the combinatorial case shows suitable resolution by obeying a weak version of the Sturm comparison principle. Furthermore the discussion in Section 2.3 demonstrated that the GFB can sometimes exhibit localization; propositions 2.3.1 and 2.3.4 showed that this phenomenon is equivalent to the existence of joint Dirichlet-Neumann spectra. Finally, in Chapter 3 we provided a peek into the intersections of harmonic analysis with high-dimensional data by engaging in an experimental discussion on the applications of diffusion maps as multi-scale frameworks that resolved finer and coarser structures within datasets and images. Observing that Diffusion Maps were a non-local framework, Neumann maps were proposed to embed subdomains of datasets. We concluded with an example which suggested that Neumann maps were more appropriate as an adaptive framework than Diffusion Maps on subdomains because they seemed to encode the ambient manifold in their embeddings.

Bibliography

[1] William K Allard, Guangliang Chen, and Mauro Maggioni. "Multi-scale geometric methods for data sets II: Geometric multi-resolution analysis". In: *Applied and Computational Harmonic Analysis* 32.3 (2012), pp. 435–462.

[2] Mikhail Belkin and Partha Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation". In: *Neural computation* 15.6 (2003), pp. 1373–1396.

[3] Türker Biyikoglu, Josef Leydold, and Peter F Stadler. *Laplacian eigenvectors of graphs: Perron-Frobenius and Faber-Krahn type theorems*. Springer, 2007.

[4] James C Bremer et al. "Diffusion wavelet packets". In: *Applied and Computational Harmonic Analysis* 21.1 (2006), pp. 95–112.

[5] Lennart Carleson. *On convergence and growth of partial sums of Fourier series*. 1966.

[6] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. 92. American Mathematical Soc., 1997.

[7] Ronald R Coifman and Matan Gavish. "Harmonic analysis of digital data bases". In: *Wavelets and Multiscale analysis*. Springer, 2011, pp. 161–197.

[8] Ronald R Coifman et al. "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps". In: *Proceedings of the national academy of sciences* 102.21 (2005), pp. 7426–7431.

[9] Ingrid Daubechies. *Ten lectures on wavelets*. Vol. 61. Siam, 1992.

[10] JB Diaz and Joyce R McLaughlin. "Sturm comparison theorems for ordinary and partial differential equations". In: *Bulletin of the American Mathematical Society* 75.2 (1969), pp. 335–339.

[11] Lawrence C Evans. *Partial differential equations*. Vol. 19. American Mathematical Soc., 2010.

[12] Michael W Frazier. *An introduction to wavelets through linear algebra*. Springer Science & Business Media, 2006.

[13] Tsogtgerel Gantumur. "Lecture Notes PDE 1, Fall 2013, McGill University". In: (2013). URL: `http://www.math.mcgill.ca/gantumur/math580f13/`.

[14] Loukas Grafakos. *Classical fourier analysis*. Vol. 2. Springer, 2008.

[15] Alexander Grossmann and Jean Morlet. "Decomposition of Hardy functions into square integrable wavelets of constant shape". In: *SIAM journal on mathematical analysis* 15.4 (1984), pp. 723–736.

[16] Alfred Haar. *Zur theorie der orthogonalen funktionensysteme*. Georg-August-Universitat, Gottingen., 1909.

[17] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. "Wavelets on graphs via spectral graph theory". In: *Applied and Computational Harmonic Analysis* 30.2 (2011), pp. 129–150.

[18] Risi Kondor, Nedelina Teneva, and Vikas Garg. "Multiresolution matrix factorization". In: *International Conference on Machine Learning*. 2014, pp. 1620–1628.

[19] Ann B Lee, Boaz Nadler, and Larry Wasserman. "Rejoinder: Treelets". In: ().

[20] László Lovász. *Large networks and graph limits*. Vol. 60. American Mathematical Soc., 2012.

[21] Stephane G Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". In: *IEEE transactions on pattern analysis and machine intelligence* 11.7 (1989), pp. 674–693.

[22] Yves Meyer. "Principe d'incertitude, bases hilbertiennes et algebres d'operateurs". In: *Séminaire Bourbaki* 662 (1985), pp. 1985–1986.

[23] Peyman Milanfar. "A tour of modern image filtering: New insights and methods, both practical and theoretical". In: *IEEE signal processing magazine* 30.1 (2012), pp. 106–128.

[24] Marıa Cristina Pereyra and Lesley A Ward. *Harmonic analysis: from Fourier to wavelets*. Vol. 63. American Mathematical Soc., 2012.

[25] Nathanaël Perraudin et al. "GSPBOX: A toolbox for signal processing on graphs". In: *ArXiv e-prints* (Aug. 2014). arXiv: 1408.5781 [cs.IT].

[26] Franz Rellich. "Darstellung der Eigenwerte von $\Delta u + \lambda u = 0$ durch ein Randintegral". In: *Math. Z.* 46 (1940), pp. 635–636.

[27] David I Shuman et al. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains". In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.

[28] Elias M Stein. "On the functions of Littlewood-Paley, Lusin, and Marcinkiewicz". In: *Transactions of the American Mathematical Society* 88.2 (1958), pp. 430–466.

[29] Robert S Strichartz. *Differential equations on fractals: a tutorial*. Princeton University Press, 2006.

[30] Vladimir M Tikhomirov. *Selected Works of AN Kolmogorov: Volume I: Mathematics and Mechanics*. Vol. 25. Springer Science & Business Media, 1991.

[31] Dengsheng Zhang. "Wavelet Transform". In: *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*. Cham: Springer International Publishing, 2019, pp. 35–44. ISBN: 978-3-030-17989-2. DOI: 10.1007/978-3-030-17989-2_3. URL: https://doi.org/10.1007/978-3-030-17989-2_3.

Appendix

5.1 Background to Chapter 2

Here we provide some intuition behind the mysterious propositions in Section 2.1. Note that proving most of these results from scratch takes a quantity of work that is significantly out of the scope of the thesis (even the dreaded Appendix!). As a consequence, the focus will be kept on techniques. First, the essential lemma of all PDE.

Lemma 5.1.1 (Du Bois-Reymond lemma). *Let $h : \Omega \mapsto \mathbb{R}$ be measurable such that*

$$\int_\Omega h\varphi \, dx = 0$$

for every $\varphi \in C_c^\infty$. Then $h = 0$ almost everywhere.

Proof. Since $\{h > 0\}$ and $\{h < 0\}$ are measurable, assume that one of them has positive measure. Without loss of generality set $E = \{h > 0\}$ and assume first that E is bounded. Then $\exists$ compact K and open V such that $K \subset E \subset V$ and K has positive measure. From Urysohn's lemma, we find a compactly supported function φ which is strictly positive on K. Then integrating h against that function produces a contradiction. □

As a corollary we have Proposition 2.1.2:

Corollary. *Let $u \in C^2$ be a Dirichlet (resp. Neumann) eigenfunction. Then u is a weak Dirichlet (resp. Neumann) eigenfunction.*

Proof. Integration by parts and the Dubois-Reymond lemma. □

Next, the intuition behind weak derivatives. The following theorem shows that ∇u is a weak derivative if and only if the finite differences $\Delta_h u = u(x+h) - u(x)$ are linearly approximated in the L^p sense by ∇u as $h \to 0$.

Theorem 5.1.2. *Let $u \in L^p(\Omega)$ where $1 < p < \infty$. Then $u \in W^{1,p} \iff ||\Delta_h f||_{L^p(\Omega_h)} = |h|||\nabla f||_{L^p(\Omega)}$ where $\Delta_h f = f(x+h) - f(x)$ and $\Omega_h = \{x \in \Omega \mid [x, x+h] \in \Omega\}$ and $[x, x+h]$ is the geodesic joining x and $x+h$.*

Proof. Here we only prove sufficiency. To see necessity, refer to [13]. Let $h = \eta t$ where $\eta \in S^{n-1}$. Then

$$|\Delta_h f| = |f(x+\eta t) - f(x)| = \int_0^t \nabla f(x+\eta s)\, ds \leqslant t^{1-1/p} \Big(\int_0^t |\nabla f(x+\eta s)|^p \, ds \Big)^{1/p}$$

Now take the pth power and integrate to get

$$\begin{aligned} ||\Delta_h f||_{L^p(\Omega_h)} &\leqslant t^{p-1} \int_{\Omega_h} \int_0^t |\nabla f(x+\eta s)|^p \, ds \, dx \\ &\leqslant t^{p-1} \int_0^t \int_{\Omega_h} |\nabla f(x)|^p \, dx \, ds = t^p ||\nabla f||_{L^p(\Omega)} \end{aligned}$$

□

The next result provides intuition behind why weak derivatives can be used to prove regularity.

Lemma 5.1.3. *Let $u \in H_0^1(\Omega)$ and $supp(u) \subseteq K$ where K is compact. Furthermore, let $f \in L^2(\Omega)$ such that*

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \text{ for every } v \in H_0^1$$

Then $u \in H^2(\Omega)$

Proof. The strategy is to explicitly calculate the weak derivative using finite differences and show that it is $O(h)$. To that end let $D_h u = u(x+h) - u(x)$. Now we use the fact that u is compactly supported by observing that $\exists K'$ compact such that $K \subseteq \text{int}(K)$ and $K + h \subseteq K'$

for small enough $h \in \mathbb{R}^n$ such that for any $v \in H_0^1$, $v(x-h) \in H_0^1$. Then we can integrate $u(x+h)$ against $v \in H_0^1$ by pullback:

$$\begin{aligned}\int_\Omega \nabla u(x+h) \cdot \nabla v \, dx &= \int_K \nabla u(x+h) \cdot \nabla v \, dx \\ &= \int_{K+h} \nabla u(x) \cdot \nabla v(x-h) \, dx \\ &= \int_\Omega \nabla u(x) \cdot \nabla v(x-h) \, dx \\ &= \int_\Omega f v(x-h) \, dx\end{aligned}$$

As a consequence we have that

$$\int_\Omega \nabla D_h u \cdot \nabla v = \int_\Omega f D_{-h} v \, dx$$

. Using the Cauchy-Schwartz inequality and the proof in Theorem 5.1.2 we have that

$$\Big| \int_\Omega \nabla D_h u \cdot \nabla v \, dx \Big| \leqslant ||f||_2 ||D_{-h} v||_2 \leqslant |h| ||f||_2 ||\nabla v||_2 = O(h)$$

Now in the above equation set $v = D_h u$. Then we have that $||\nabla D_h u||_2 \leqslant |h| ||f||_2$. In particular, we have that

$$||D_{te_i} \partial_i u||_2 = O(h)$$

so from Theorem 5.1.2 $\partial_i u$ is H^1 for any i so $u \in H^2$. □

We also state the Sobolev Inequalities which provide a bridge between the weak and strong formulations:

Theorem 5.1.4 (Gagliardo-Nirenberg-Sobolev Inequalities). *Let $u \in W_0^{1,p}(\Omega)$. Then the following inequalities hold:*

$$||u||_q \leqslant C(n,p) ||\nabla u||_p, \qquad 1 \leqslant q < \frac{np}{n-p}, \tag{5.1.1}$$

$$\sup_\Omega |u| \leqslant C(n,p,\Omega) ||\nabla u||_p, \qquad p > n. \tag{5.1.2}$$

The GNS inequalities show that one can bound the norm of a function in $W_0^{1,p}(\Omega)$ if one knows that the norm of its derivative is finite. By the uniformity of the constant in Equation

5.1.1 we have that $W_0^{1,p} \subseteq L^q$. This result is known as the **Sobolev Embedding Theorem**. Furthermore, if the degree of control on the derivative is larger than the dimension, then the function must be continuous with an explicit bound on the sup norm.

Lastly, we provide a way to connect analyticity with weak derivatives:

Lemma 5.1.5. *Let $f : \Omega \mapsto \mathbb{R}$ and suppose for every y there exists $r, \delta > 0$ and $M < \infty$ such that for every $x \in B_r(y)$ the following estimate holds*

$$||f||_{C^k} \leqslant M \frac{k!}{\delta^k}$$

Then f is analytic

Proof. The proof follows by converting the multivariable function to a single variable function and then computing the Taylor series of the single variable function. To that end fix $y \in \Omega$ and let δ, M, r be as given in the statement. Without losing generality suppose $y = 0$. Let $x \in B_r(y)$ and set $g(t) = f(xt)$. Now observe that via the chain rule $g^{(k)}(x) = (x_1\partial_1 + \dots x_n\partial_n)^k f(xt) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha \partial^\alpha f(xt)$ due to the multinomial theorem and the chain rule. Finally, we observe that $g(1) = f(x)$ and the Taylor series with remainder states that

$$g(1) = \sum_{k=0}^{m-1} \frac{g^{(k)}(0)}{k!} + \frac{g^{(k+1)}(s)}{(k+1)!}$$

where $0 < s < 1$. We plug in the formula for $g^{(k)}(0)$ in terms of f and notice that the remainder term decays to 0 as $k \to \infty$ because of the decay estimate. Thus the Taylor series converges uniformly to f for every $y \in \Omega$ so f is analytic. □

Lemma 5.1.6. *Let $f : \Omega \mapsto \mathbb{R}$ and suppose for every y there exists $r, \delta > 0$ and $M < \infty$ such that for every $x \in B_r(y)$ the following estimate holds*

$$||f||_{H^k} \leqslant M \frac{k!}{\delta^k}$$

Then f is analytic.

Proof. The proof is a consequence of the generalized Sobolev inequality which states that if $u \in W^{k,2} = H^k$ where $k > n/2$ then

$$||u||_{C^{k-\lfloor \frac{n}{2} \rfloor + 1}} \leqslant C||u||_{H^k}$$

Since $f \in H^k$ for every k in the ball $B_r(y)$, we have that $||f||_{C^k} \leqslant C||f||_{H^{k+p}} \leqslant cM\frac{(k+p)!}{\delta^{k+p}}$ where $p = \left\lfloor \frac{n}{2} \right\rfloor + 1$. Now all we need to do is make the bound look like the condition of Lemma 5.1.5. To do so, when $k \leqslant p$ we have that $(k+p)! \leqslant (2p)! \leqslant 2^p p^{2p}$. Furthermore, when $k > p$ from Stirling's Formula we have that $(k+p)! \leqslant k!(2k)^p \leqslant p^p e^k k!$. Plugging these estimates into the estimate for $||f||_{C^k}$ we get that $||f||_{C^k} \leqslant Ck!(\delta')^{-k}$ where the constants are absorbed into C and $\delta' = (\delta/e)$. □

The road to the analytic regularity theorem is as follows: Using arguments similar to 5.1.3 show that when $f \in H^k$, $u \in H^{k+2}$. This is known as the Interior Regularity Theorem [11]. Next, combine Theorem 5.1.4 with the fact that $f = 0$ to get that u is smooth. Lastly, use the bounds from the Interior Regularity Theorem with 5.1.6 to get that u is analytic. Note that all these proofs make heavy use of the du Bois-Reymond lemma.

5.2 Code Listing

5.2.1 Contents

5.2.2 Utilities

In this section we describe all the helper functions required for the simulations.

- Generating Haar wavelets of given size

```
function [Hr]=generate_haar(N)
% Author: Kamlesh Pawar
% Input :
%     N : size of matrix to be generated, N must be some power of 2.
```

```
% Output:
%    Hr : Haar matrix of size NxN

if (N<2 || (log2(N)-floor(log2(N)))~=0)
    error('The input argument should be of form 2^k');
end

p=[0 0];
q=[0 1];
n=nextpow2(N);

for i=1:n-1
    p=[p i*ones(1,2^i)];
    t=1:(2^i);
    q=[q t];
end
Hr=zeros(N,N);
Hr(1,:)=1;
for i=2:N;
    P=p(1,i); Q=q(1,i);
    for j= (N*(Q-1)/(2^P)):(N*((Q-0.5)/(2^P))-1)
        Hr(i,j+1)=2^(P/2);
    end
    for j= (N*((Q-0.5)/(2^P))):(N*(Q/(2^P))-1)
        Hr(i,j+1)=-(2^(P/2));
    end
end
Hr=Hr*(1/sqrt(N));
end
```

- Graph Data Structures Converting a graph from `Graph` type to `struct`

```
function [G] = graph2struct(H)
% Takes a graph type and outputs a graph struct to use for
% the gsp box
A = adjacency(H, 'weighted');
N = max(size(A));
G.N = N;
G.W = A;
% Assign a default circular embedding
G.coords=[(cos((0:N-1)*(2*pi)/N))',(sin((0:N-1)*(2*pi)/N))'];
G.plotting.limits=[-1,1,-1,1];
G = gsp_graph_default_parameters(G);
```

- Creating a random graph

```
function A = RandomGraph(N,p)
A = zeros(N,N);
for i = 1:N
    for j = i:N
        A(i,j) = binornd(1,p);
        A(j,i) = A(i,j);
    end
end
A = A - diag(diag(A));

end
```

- Creating an N-Fan graph

```
function A = nfan(N)
%Return an N fan graph with 2n+1 vertices
A = zeros(2*N+1);
```

```
A(1,:) = ones(1,2*N+1); %The first vertex is the center of the fan
A(1,1) = 0;
for i = 1:N
    %i loops through each blade
    A(2*i,1) = 1;
    A(2*i + 1, 1) = 1;
    A(2*i, 2*i + 1) = 1;
    A(2*i + 1, 2*i) = 1;
end

end
```

- Graph Drawing

  Computing the circular embedding of an N-vertex graph

```
function E = embedding(N,k)
% N is the number of points
% k is the center of the circle
n = 0:1:N-1;
x = cos(n*(2*pi/N)) + k(1);
y = sin(n*(2*pi/N)) + k(2);
E = [x;y]
end
```

- Subgraphs Finding the boundary from the index set:

```
% This function gives the boundary of a subgraph indexed by sub

% Arguments:
% 1. G is the ambient graph encoded as a struct
% 2. sub indexes the subset
```

```
function Bindex = boundary(G, sub)

if ~isstruct(G)
   G = graph2struct(G);
end

A = G.W; % Computing the adjacency matrix of G
Bindex = zeros(length(sub),length(A)); % Stores the index values
for i=1:length(sub)
    for j=1:length(A(1,:))
        if A(sub(i),j) ~= 0 && isempty(sub(sub==j))
            Bindex(i,j) = j;
        end
    end
end

Bindex = unique(Bindex(:));
Bindex = Bindex(Bindex ~=0);

end
```

Computing subgraphs

```
% This function gives three important subgraphs:
% H -- Subgraph of G induced by sub
% I -- Subgraph of G induced by sub with boundary edges
% J -- Subgraph of G induced by sub union boundary vertices
% deltaS -- the boundary
% G -- the graph
% sub -- selection of the vertices

% First some mopping up. If G isn't a struct, we convert it to a struct
% object. If it is then we keep it.
```

```
function [H,I,J, deltaS] = subs(G,sub)

if ~isstruct(G)
   G = graph2struct(G);
end
% A stores the adjacency (or weight) matrix

A = G.W;

% find boundary vertices

deltaS = boundary(G,sub);

% find induced subgraph

H = gsp_subgraph(G,sub);

% graph with subgraph and boundary

AdjS = zeros(length(sub)+length(deltaS)); % Creating an adjacency matrix
AdjS(1:length(sub), 1:length(sub)) = A(sub,sub);
AdjS(length(sub)+1:end,1:length(sub)) = A(deltaS, sub);
AdjS(1:length(sub), length(sub)+1:end) = (A(deltaS, sub))' ;
I = graph(AdjS);
I = graph2struct(I);

I = gsp_subgraph(G,[sub';deltaS]);
I.W(end-length(deltaS)+1:end,end-length(deltaS)+1:end) = zeros(length(deltaS));

% % graph induced by subgraph union boundary

J = subg(G,[sub';deltaS]);
```

```
% AdjS_deltaS = A([sub reshape(deltaS, [1,length(deltaS)])], [sub reshape(deltaS, [
% % Creating reordered adjacency with subgraph vertices first
% J = graph(AdjS_deltaS);
% J = graph2struct(J);
%
```

- Image Processing Computing the graph of an image:

```
function [G] = im2graph(I)
%takes an image as a matrix
%outputs a graph as a struct with lattice coords

[n,m] = size(I);
W = zeros(n*m);
sigma_p = max(max(I));
coordinates = zeros(n*m,2);
for i=1:n*m
       row = floor((i-1)/n) + 1;
       column = mod((i-1),m) + 1;
       coordinates(i,2) = 1 - (row/n) + (1/n);
       coordinates(i,1) = (column/m) - (1/m);
end

for i = 1:n*m
    for j = 1:n*m
           row1 = floor((i-1)/n) + 1; %row position of the ith pixel
           column1 = mod((i-1),m) + 1; %column position of the ith pixel
           row2 = floor((j-1)/n) + 1; %row position of the ith pixel
           column2 = mod((j-1),m) + 1; %column position of the ith pixel
           g_distance = norm(coordinates(i,:)-coordinates(j,:));
           p_distance = I(row1, column1) - I(row2,column2);
```

```
            W(i,j) = exp(-((g_distance)^2)/(2))*exp(-((p_distance)^2)/(2*sigma_p));
    end
end
G.W = sparse(W);
G.coords = coordinates;
G.plotting.limits=[0,1,0,1];
G = gsp_graph_default_parameters(G);
```

- Nodal Sets

  Computing Nodal edges:

```
function E = crossings(A,f)
E = 0;

N = max(size(A));
      for i = 1:N
          for j =i:N
              if (f(i)*f(j) \ensuremath{<} 0) \&\& (A(i,j) == 1)
                 E = E+1;
              end
          end
      end
end
```

  Computing $G(\lambda)$ from Proposition 2.2.1:

```
function G = lowerbound(A, lambda)

N = max(size(A));
G = 0 ;
   for i = 1:N
       if sum(A(i,:)) < lambda
```

```
                G = G + 1 ;
            end
        end
```

5.2.3 General Graph Theory

- Simulations with subgraphs and their boundaries

```
% Compute the graph here. Some sample computations are given below

% Erdos Renyi random graph

e = 50;
v = 100;
G := Graph::createRandomGraph(v,e, undirected):

% Random Graph with a bernoulli distributed adjacency matrix
p=0.05;
N = 100;
A = RandomGraph(N,p);
G = graph(A);
plot(G)

% Path graphs

vec =  zeros(1,10);
vec(2) = 1;
Adj = toeplitz(vec);
G = graph(Adj)
plot(G)

% Cyclic graph

vec = zeros(1,10);
vec(2) = 1; vec(length(vec)) = 1;
```

```
Adj = toeplitz(vec);
G = graph(Adj);
plot(G)

% Complete graph
weights = [1 -2 3 -4 5 -6 7 -8 9];
vec = ones(1,9);
vec(1) = 0;
Adj = toeplitz(vec);

% Define the graph subset here in row or column vector form

sub = randperm(100,20);

create the three important subgraphs
[S, S_deltaS, S_UdeltaS, deltaS] = subs(G,sub);

define new colormap
mymap = [1 0 0
         0 0 0
         0 0 1];
colormap(mymap);

% plotting subgraph with boundary

colours = zeros(1,length(sub)+length(deltaS));
colours(1:length(sub)) = 1;
colours(length(sub)+1:end) = 2;

% subroutine for assigning colours to edges

Edges = table2array(S_deltaS.Edges);
Edges = Edges(:,1:2); % Extract the set of edges denoted by ordered pairs
e_colours = zeros(1,length(Edges)); % Set of edges
```

```
for i = 1:length(e_colours)
    if Edges(i,1) <= length(sub) && Edges(i,2) <= length(sub)
        e_colours(i) = 1; % case when the edge is in S
    else
        e_colours(i) = 2;
    end
end

% end of subroutine

p = plot(S_deltaS, 'MarkerSize', 10, 'LineWidth', 2);
legend('Subgraph', 'Boundary');
p.NodeCData = colours;
p.NodeLabel = [];
p.EdgeCData = e_colours;
p

% plotting graph induced by subgraph union boundary

v_colours = zeros(1,length(sub)+length(deltaS));
v_colours(1:length(sub)) = 2;
v_colours(length(sub)+1:end) = 3;

% % subroutine for assigning colours to edges

Edges = table2array(S_UdeltaS.Edges);
Edges = Edges(:,1:2); % Extract the set of edges denoted by ordered pairs
e_colours = zeros(1,length(Edges)); % Set of edges
for i = 1:length(e_colours)
    if Edges(i,1) <= length(sub) && Edges(i,2) <= length(sub)
        e_colours(i) = 2; % case when the edge is in S
    elseif Edges(i,1) > length(sub) && Edges(i,1) > length(sub)
        e_colours(i) = 3; % case when the edge is not in S or the boundary
```

```
    else
        e_colours(i) = 1;

    end
end

% % end of subroutine

subgraph_location = embedding(length(sub),[0 0]);
boundary_location = embedding(length(deltaS),[4 0]);
location = [subgraph_location boundary_location];
p = plot(S_UdeltaS, 'MarkerSize', 10, 'LineWidth', 2);
legend('Subgraph', 'Boundary');
p.NodeCData = v_colours;
p.NodeLabel = [];
p.EdgeCData = e_colours;
p.XData = location(1,:);
p.YData = location(2,:);
p
```

5.2.4 Fourier Analysis on graphs

- Computing Neumann and Dirichlet Operators

```
function [N,D, B, deltaT_S, T_S, N_mat] = Neumann_Dirichlet(G,sub)
% Computes the Neumann and dirichlet operators
% operator on the subgraph induced by sub
% G is the graph. Could be in struct type.
% sub is the selection of vertices

if ~isstruct(G)
   G = graph2struct(G);
end
```

```
[~, S_deltaS, S_UdeltaS, ~] = subs(G,sub);
L = S_UdeltaS.L; % note that L is stored as sparse double
D = L(1:length(sub), 1:length(sub)); % the dirichlet matrix
B = -L(length(sub)+1:end,1:length(sub)); %the boundary map
l = S_deltaS.L;
deltaT_S = l(length(sub)+1:end, length(sub)+1:end);
N = sparse(D - (B')*(diag(1./diag(deltaT_S)))*B);
diagonal = diag(l);
T_S = diag(diagonal(1:length(sub)));
N_mat = vertcat(sparse(eye(length(sub))), sparse((diag(1./diag(deltaT_S)))*B));

end
```

- Combinatorial Laplacian Spectra of some common graphs

```
N = 10;
A = RandomGraph(N);
G = graph(A);
plot(G, 'NodeColor', 'r', 'MarkerSize', 10, 'LineWidth', 2);

% Path graphs

% vec =  zeros(1,10);
% vec(2) = 1;
% Adj = toeplitz(vec);
% G = graph(Adj)
% plot(G)

% Cyclic graph
% S
% vec = zeros(1,10);
% vec(2) = 1; vec(length(vec)) = 1;
% Adj = toeplitz(vec);
```

```
% G = graph(Adj);
% plot(G)

% Complete graph
%weights = [1 -2 3 -4 5 -6 7 -8 9];
% vec = ones(1,9);
% vec(1) = 0;
% Adj = toeplitz(vec);

% Emb = embedding(N);
% G = graph(A);
% L = full(laplacian(G));
% [V, D] = eig(L);
%plot(G, 'XData', Emb(1,:), 'YData', Emb(2,:), 'ZData', V(:,6));
%stem(1:N, V(:,9))
```

- Exploring $|\mathcal{E}|$ vs $G(\lambda)$

```
N = 20;
figure
title('Comparing $$\mathcal{E}$$ to G');
for k = 1:9
    A = RandomGraph(N);
    G = graph(A);
    L = full(laplacian(G));
    [V, D] = eig(L);
    %eigs stores the eigenvalues
    eigs = diag(D);
    %C stores the edge crossing number
    C = zeros(1,N);
    for i = 1:N
        C(i) = crossings(A, V(:,i));
```

```
    end
    %Creating the G vector. For some reason Matlab isn't allowing me to broad
    %cast so I'll just do it via a for loop. Sigh.
    G = zeros(1,N);
    for i = 1:N
        G(i) = lowerbound(A, eigs(i));
    end
    subplot(3,3,k)
    plot(eigs(2:N), C(2:N), 'r.');
    hold on
    plot(eigs(2:N), G(2:N)/2, 'b-');
end
```

- Laplacian Spectra of Fans

```
A = nfan(3);
G = graph(A);
D = diag(A*ones(max(size(A)),1));
plot(G);
L = full(laplacian(G));
L_norm = D^(-1/2)*L*D^(-1/2);
[V,D] = eig(L_norm);
```

5.2.5 Diffusion maps

- Computing a Laplacian Eigenmap

```
function [Diff_maps] = My_Eigenmaps(G,t,dim)
% Spits out 5 dimensional laplacian eigenmaps given a graph G in struct
% form
N = max(size(G.W));
% compute the random walk matrix here
% t is the scaling factor in diffusion map
```

```
lap = gsp_create_laplacian(G, 'normalized'); %lap will be a struct
D = diag(full(lap.W)*(ones(N,1))); %Stores diagonal matrix
M = eye(N) - full(lap.L); %Regularized random walk
[X, Lambda] = eigs(M,dim+1,'largestabs'); %Compute spectral decomposition up to 5 eigenvectors
Phi = D^(1/2)*X; %Phi matrix
Psi = D^(-1/2)*X; %Psi matrix; you want to extract its columns!
Diff_maps = (Psi)*(Lambda^t); %Multiplying each column with the respective eigenvalue
Diff_maps = Diff_maps(:,2:end); %Dropping the first column as it's all a constant
% Each column of diff_maps contains a coordinate of the diffusion map
% to plot it, plot Diff_maps(:,j) in the jth coordinate
end
```

- Laplacian eigenmaps with digital weights

```
First we provide a graph G. Enter graph here
N = 128; % Number of vertices
G = gsp_spiral(N,3); % Creating a graph in struct version

Diff_maps = My_Eigenmaps(G,1);

% Plotting the jdim diffusion map
%plot(Diff_maps(:,1),Diff_maps(:,2),'ro');
plot3(Diff_maps(:,3),Diff_maps(:,4),Diff_maps(:,1),'ro');
```

- Diffusion maps for ring, spiral, swiss roll, sphere, and stochastic block graphs

```
% Uncomment each section to visualize the respective laplacian eigenmap

% Map for rings

G = gsp_ring(1000);
W = full(G.W);
[mappedX, mapping, lambda] = lapbasic(W, 3, 1, 'JDQR');
for i=0:0.1:1
```

```
    plot(mappedX(:,1),mappedX(:,3))
    hold on
end

% % Map for spiral
%
% G = gsp_spiral(100,3);
% W = full(G.W);
% [mappedX, mapping, lambda] = lapbasic(W, 3, 1, 'JDQR');
% for i=0:0.1:1
%      plot3(((lambda(1))^(-i))*mappedX(:,1),((lambda(2))^(-i))*mappedX(:,2),((lambda(3)
%      hold on
% end

% % Map for sphere
%
% G = gsp_sphere(100);
% W = full(G.W);
% [mappedX, mapping, lambda] = lapbasic(W, 3, 1, 'JDQR');
% for i=0:0.1:1
%     plot3(((lambda(1))^(-i))*mappedX(:,1),((lambda(2))^(-i))*mappedX(:,2),((lambda(3))
%     hold on
% end

% Map for swiss roll

% G = gsp_swiss_roll(500);
% W = full(G.W);
% [mappedX, mapping, lambda] = lapbasic(W, 3, 1, 'JDQR');
% % for i=0:0.1:1
% %     plot(((lambda(1))^(-i))*mappedX(:,1),((lambda(2))^(-i))*mappedX(:,2))
% %     hold on
```

```
% % end
% plot3(((lambda(1))^(-i))*mappedX(:,1),((lambda(2))^(-i))*mappedX(:,2),((lambda(3))^(-i))*mapp

% Map for stochastic block graphs

% G = gsp_stochastic_block_graph(1024,10);
% W = full(G.W);
% [mappedX, mapping, lambda] = lapbasic(W, 3, 1, 'JDQR');
% for i=0:0.1:1
%     plot3(((lambda(1))^(-i))*mappedX(:,1),((lambda(2))^(-i))*mappedX(:,2),((lambda(3))^(-i))*
%     hold on
% end
```

5.2.6 Figures

Chapter 1

- Figure 1.1

```
H = generate_haar(512);
times = linspace(0,1,512);
figure;
for i=1:8
    if i==1
        titlestring = strcat("$\varphi$");
        minlim = -1;
        maxlim = 1;
    else
         j = floor(log2(i-1));
         k = (i-1)-2^j;
         titlestring = strcat("$\psi_{",num2str(j),",",num2str(k),"}$");
         maxlim = max(H(i,:));
         minlim = min(H(i,:));
```

```
    end
    subplot(2,4,i);
    set(gca,'TickLabelInterpreter','latex');
    set(groot, 'DefaultLegendInterpreter','latex');
    plot(times,H(i,:),'LineWidth',2,'DisplayName','Level 8 Haar approximation');
    yticks([]);
    xticks([0 0.25 0.50 0.75 1]);
    xticklabels({'0', '$\frac{1}{4}$', '$\frac{1}{2}$', '$\frac{3}{4}$', '1'});
    set(gca,'FontSize',16);
    legend('FontSize',11);
    title(titlestring,'interpreter','latex','FontSize',20);
```

- Figure 1.2

```
load BabyECGData;
    % figure;
    % p1 = plot(times,HR,'-');
    % xlabel('Hours');
    % ylabel('Heart Rate');
    % p1.Color(4) = 0.25;
    % hold on;
    [a,d] = haart(HR,'integer');
    % HaarHR = ihaart(a,d,1,'integer');
    % plot(times,HaarHR,'Linewidth',1)
    % title('Haar Approximation of Heart Rate')
    imz = zeros(10,2048);

for i = 1:10
        HaarHR = ihaart(a,d,i,'integer');
        imz(i,:) = HaarHR';
    end
```

```
    figure;
    subplot(2,1,1);
    set(gca,'TickLabelInterpreter','latex');
    set(groot, 'DefaultLegendInterpreter','latex');
    p1 = plot(times,HR,'-','DisplayName', 'Heart Rate');
    xlabel('Hours', 'interpreter','latex', 'FontSize',16);
    ylabel('Heart Rate','interpreter', 'latex','FontSize',16);
    p1.Color(4) = 0.2;
    hold on
    p2 = plot(times,imz(3,:),'r-','LineWidth',1,'DisplayName','Level 8 Haar approximation'
    hold on
    p3 = plot(times,imz(7,:),'k--','LineWidth',2,'DisplayName','Level 4 Haar approximation
    legend('FontSize',11);

subplot(2,1,2);
    colormap copper
    set(gca,'TickLabelInterpreter','latex');
    set(groot, 'DefaultLegendInterpreter','latex');
    image(imz,'CDataMapping', 'scaled');
    cbh = colorbar;
    cbh.Ticks = [];
    ylabel(cbh, 'Heart Rate','interpreter','latex','FontSize',16);
    xticks([]);
    %xlabel("Hours", 'interpreter','latex','FontSize', 16);
    ylabel("Scale (j)", 'interpreter','latex','FontSize', 16);
```

- Figure 1.3

```
load mandrill
%im = imread('Rcirc.png');
im = imresize(X,[512 512]);
```

```
%im = im(:,:,1);

[a,h,v,d] = haart2(im,'integer');
figure;

for i=0:8
    row = floor(i/3) + 1;
    column = mod(i,3) + 1;
    subplot(3,3,i+1);
    imrec = ihaart2(a,h,v,d,i,'integer');
    colormap parula
    imagesc(imrec);
    title(strcat('Level', " ", num2str(8-i + 1)), 'interpreter', 'latex','FontSize'
    axis off;
end

% to extract a 2^N x 2^N sized image, just pick d(N:end) and run ihaart
```

- Figure 1.5

```
im = imread('obama.jpg');
im = imresize(im, [512 512]);
[a,h,v,d] = haart2(im,'integer');
D = d(5:end);
H = h(5:end);
V = v(5:end);
Imz = ihaart2(a,H,V,D,1,'integer');
Imz = double(Imz(:,:,1));
G  = im2graph(Imz);
S = gsp_compute_fourier_basis(G);
U = full(S.U);
```

```
fiedler_vector = U(:,2);
M = median(fiedler_vector);
classifier = fiedler_vector > 0;
gsp_plot_signal(G,classifier);
colormap flag
```

Chapter 2

- Comparing the DHT and DFT of a 32-bit image with a central spot

```
Img = zeros(32,32);
Img(16,16) = 1;
Img(16,17) = 1;
Img(17,16) = 1;
Img(17,17) = 1;

FT = fft2(Img);
[a,h,v,d] = haart2(Img, 'integer');

HT = cell2mat(d(1));

figure();

subplot(2,2,1);
imagesc(Img);
caxis('manual');
caxis([-1 1])
title('Original Image','interpreter','latex','FontSize',20);
set(gca,'XColor', 'none','YColor','none')

subplot(2,2,2);
imagesc(HT);
caxis('manual');
```

```
caxis([-1 1])
title('16-point Haar Coefficients','interpreter','latex','FontSize',20);
set(gca,'XColor', 'none','YColor','none')

subplot(2,2,3);
imagesc(real(FT));
caxis('manual');
caxis([-1 1])
title('Real part of Fourier coefficients','interpreter','latex','FontSize',20);
set(gca,'XColor', 'none','YColor','none')

subplot(2,2,4);
imagesc(imag(FT));
caxis('manual');
caxis([-1 1])
title('Imaginary part of Fourier coefficients','interpreter','latex','FontSize',20)
set(gca,'XColor', 'none','YColor','none');
cbh = colorbar;
cbh.Ticks = [-1 1];
ylabel(cbh, 'Luminescence/Coefficient Value','interpreter','latex','FontSize',20);
```

Chapter 3

- Figure 3.1

```
Let's create a circulant matrix!

i = 10;
v = [0 ones(1,i) zeros(1,99-(2*i)) ones(1,i)];

A = toeplitz([v(1) fliplr(v(2:end))], v);

P = eye(100);
P = P(randperm(100),:);
```

```
Adj = P*A*P';

G = graph(Adj);

G = graph2struct(G);

Diff_Maps = My_Eigenmaps(G,1,3);

% Plotting the jdim diffusion map
subplot(1,2,1)
colormap gray
imagesc(Adj);
title("Adjacency Matrix of a circulant graph $V=100$, $k=20$",'fontsize',16,'interpreter',

subplot(1,2,2);
scatter(Diff_Maps(:,1),Diff_Maps(:,2));
title("2 Dimensional Diffusion embedding",'fontsize',16,'interpreter','latex');
%plot3(Diff_maps(:,1),Diff_maps(:,2),Diff_maps(:,3),'ro');
```

- Figure 3.2

```
G = gsp_torus(32,32);

Diff_Maps = My_Eigenmaps(G,1,5);

subplot(1,2,1)
plot3(G.coords(:,1),G.coords(:,2),G.coords(:,3),'ro');
title("Uniformly Sampled points on a Torus",'fontsize',18,'interpreter','latex');

subplot(1,2,2);
plot3(Diff_Maps(:,1),Diff_Maps(:,2),Diff_Maps(:,3),'bo-');
title("3 Dimensional Diffusion embedding",'fontsize',18,'interpreter','latex');
%plot3(Diff_maps(:,1),Diff_maps(:,2),Diff_maps(:,3),'ro');
```

- Figure 3.3

```
% Extracting the image
Iz = imread('Rcirc.png');
im = imresize(Iz,[32 32]);
im = im(:,:,1);
H = im2graph(im2double(im));
Diff_maps = My_Eigenmaps(H,1,5);
figure;
subplot(2,2,1);
imagesc(Iz(:,:,1));
title('Original Image','interpreter','latex','FontSize',20);
set(gca,'xtick',[]);
set(gca,'ytick',[]);
```

- Figure 3.4

```
Iz = imread('Rcirc.png');
im = imresize(Iz,[32 32]);
im = im(:,:,1);
H = im2graph(im2double(im));
t = [0 0.25 0.5 0.75 1];
 for i=t
     Diff_maps = My_Eigenmaps(H,i,5);
     scale = num2str(i);
     labelstring = strcat('t=',scale);
     scatter3(Diff_maps(:,1),Diff_maps(:,2),Diff_maps(:,3),'DisplayName',labelstrin
     hold on;
 end

legend;
```

```
scatter(Diff\_maps(:,1),Diff\_maps(:,2),pointsize,Diff\_maps(:,3))
pointsize = 20; colorbar jet;
scatter(Diff\_maps(:,1),Diff\_maps(:,2),pointsize,Diff\_maps(:,3)); colormap jet;
```

- Figure 3.5a

```
N = 256;
G = gsp_ring(N);

sub = N/4:(3*N/4);

NDiff_Maps = Neumann_DiffMaps(G,sub,3,0.1);

[S, S_deltaS, S_UdeltaS, deltaS] = subs(G,sub);

Diff_Maps = gsp_laplacian_eigenmaps(S,3);

figure;
subplot(2,2,1);
gsp_plot_graph(G);
title('Original Graph','interpreter','latex','FontSize',16);

subplot(2,2,2);
gsp_plot_graph(S);
title('Subgraph','interpreter','latex','FontSize',16);

subplot(2,2,3);
gsp_plot_graph(S_deltaS);
title('Subgraph with Boundary','interpreter','latex','FontSize',16);

subplot(2,2,4);
plot3(NDiff_Maps(:,3),NDiff_Maps(:,2),NDiff_Maps(:,1),'ro','DisplayName','Neumann');
```

```
hold on;
plot3(Diff_Maps(:,3),Diff_Maps(:,2),Diff_Maps(:,1),'bo','DisplayName','Diffusion');
title('Diffusion Embeddings','interpreter','latex','FontSize',16);
legend;
```

- Figure 3.5a

```
N = 256;
G = gsp_spiral(N);

sub = 1:floor(N/3);

NDiff_Maps = Neumann_DiffMaps(G,sub,3,0.1);

[S, S_deltaS, S_UdeltaS, deltaS] = subs(G,sub);

Diff_Maps = gsp_laplacian_eigenmaps(S,3);

figure;
subplot(2,2,1);
gsp_plot_graph(G);
title('Original Graph','interpreter','latex','FontSize',16);

subplot(2,2,2);
gsp_plot_graph(S);
title('Subgraph','interpreter','latex','FontSize',16);

subplot(2,2,3);
gsp_plot_graph(S_deltaS);
title('Subgraph with Boundary','interpreter','latex','FontSize',16);
```

```
subplot(2,2,4);
plot3(NDiff_Maps(:,3),NDiff_Maps(:,2),NDiff_Maps(:,1),'ro','DisplayName','Neumann');
hold on;
plot3(Diff_Maps(:,3),Diff_Maps(:,2),Diff_Maps(:,1),'bo','DisplayName','Diffusion');
title('Diffusion Embeddings','interpreter','latex','FontSize',16);
legend;
```

- Figure 3.5b

```
% Experiments with spheres
size = 512;
sphere_graph = gsp_sphere(size);
coordinates = sphere_graph.coords;
elevation = coordinates(:,3);
polarcap = coordinates(elevation > 1/2,:);
%plot3(polarcap(:,1),polarcap(:,2),polarcap(:,3),'ro');
cap = find(elevation > 1/2);
distances = gsp_distanz(coordinates',coordinates');
eps=0.5;
weightmatrix = exp(-(1/(2*(eps)^2))*(distances.^2)) - eye(size);
S = graph(weightmatrix,'upper');
S = graph2struct(S);
S.coords = coordinates;

% Run Neumann Diffusion on the polar cap
NDiff_Maps = Neumann_DiffMaps(S,cap',5,1);

% Run Standard Diffusion on the polar cap
[T, T_deltaT, T_UdeltaT, deltaT] = subs(S,cap');
Diff_maps = My_Eigenmaps(T,1,5);

%plot both
```

```
figure;
subplot(2,2,1);
plot3(coordinates(:,1),coordinates(:,2),coordinates(:,3),'bo','DisplayName','Sphere
hold on;
plot3(coordinates(cap,1), coordinates(cap,2), coordinates(cap,3),'ro','DisplayName'
title("Sphere and Polar Cap",'interpreter','latex','FontSize',16);
legend;

subplot(2,2,2);
plot3(Diff_Maps(:,3),Diff_Maps(:,2),Diff_Maps(:,1),'ro');
title('Diffusion Map','interpreter','latex','FontSize',16);

subplot(2,2,3);
plot(NDiff_Maps(:,1),NDiff_Maps(:,2),'bo');
title('2-D Neumann Map','interpreter','latex','FontSize',16);

subplot(2,2,4);
plot3(NDiff_Maps(:,1),NDiff_Maps(:,2),NDiff_Maps(:,3),'bo');
title('3-D Neumann Map','interpreter','latex','FontSize',16);
```

Corrections

When originally submitted, this honours thesis contained a number of typographical and technical errors. The author is grateful for their readers for having spotted these errors and suggested stylistic changes whenever needed, for example, to highlight original contributions. The current version includes these changes. The following is the list of all corrections:

Various places in the thesis. There were 15 syntactical errors involving grammatical, typographical, or punctuation mistakes pointed out by the readers of the original version. These have all been corrected.

Figures. All Figures except 3.2, 3.4, 3.5b, and 3.5a were remade for better readability. Code for generating these figures can be found in subsection 5.2.6 of the Appendix. The captions were altered to be more descriptive of their respective figures.

References. References were added to Examples 1.2.1 and 1.2.3, and Theorems 1.1.1, 1.1.2, 1.1.3, 2.1.12, and 2.1.5.

Other changes:

Title of Section 1.1: "Basics" was replaced with "Fundamentals."

p.2, l.14: Added here from Section 1.4: "Lastly we add [...] a fruitful theory of a multiresolution framework." "Lastly we add that" was replaced with "Consequently."

p.2, l.21: Added for transition: "But we first revisit [...] the following section."

p.4, l.1: Replaced $=$ with $\underset{\text{a.e}}{=}$.

p.4, l.2: Replaced $\mathbb{S}^1$ with $[-\pi, \pi]$.

p.4, l.2: Added $\frac{1}{2\pi}$ to the left-hand side of Equation 1.1.4.

p.4, l.9: Replaced "$L^1(\mathbb{R})$" with "$L^1(\mathbb{R}) \bigcap L^2(\mathbb{R})$." **p.9, l.10**: Changed subscript from "j" to "n."

p.9, l.10: Replaced "$\psi(2^j x - k)$" with "$\psi_{j,k}$."

p.11, l.17: Added a reference to the implementation of the 2-D Discrete Haar Wavelet transform for more information on Figure 1.3.

p.14, l.3: Added "and σ_x and σ_p are position and luminescence sensitivities respectively."

p.15, l.4: Replaced the previous sentence at this location with one emphasizing the novel result: "We show [...] do not exist on Euclidean domains."

p.15, l.12: The following sentences were added to emphasize the novel results: "All the results mentioned in this section are novel; [...] The code can be found in the Appendix at the end of this work."

p.18, l.1: Added "open" after "...any."

p.18, l.11: Added "We will occasionally drop Ω when there is no ambiguity about the domain."

p.18, l.17: Added "For the remainder of this section, we assume $\Omega \subseteq \mathbb{R}^n$ is an open simply connected domain."

p.19, l.4: Corrected Definition 2.1.3 to include weak partial derivatives.

p.20, l.13: Added a blank space before "...$\exists$."

p.21, l.7: Moved $||u||_X$ to the right hand side of the inequality.

p.21, l.19: Removed "separable" and added "When H is separable, i.e it admits a countable basis."

p.26, l.19: Changed w_i to d_i.

p.28: Replaced all instances of the subset A with F.

p.28: Changed all instances of "D" to "T."

p.29, l.16: Replaced "crossing" with "nodal."

p.29: Changed all instances of "W_{ij}" to "w_{ij}."

p.33, l.3: Removed "the ambient graph."

Chapter 3: Replaced all instances of "D" denoting the degree matrix with "T."

p.44, l.2: Added "(a graph whose adjacency matrix is circulant)."

p.44, l.8: Added for clarity: "A set of points [...] angle with the center)."

Appendix: A listing of all of the author's code was added to the Appendix.