TOPICS IN HARMONIC ANALYSIS ON COMBINATORIAL GRAPHS

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Getnet Abebe Gidelew
May, 2014

Examining Committee Members:

Isaac Pesenson, Advisory Chair, Mathematics
Shiferaw Berhanu, Mathematics
Gerardo Mendoza, Mathematics
Krzysztof Nowak, External Member, Drexel University

ABSTRACT

TOPICS IN HARMONIC ANALYSIS ON COMBINATORIAL GRAPHS

Getnet Abebe Gidelew
DOCTOR OF PHILOSOPHY

May, 2014
Professor Isaac Pesenson, Chair

In recent years harmonic analysis on combinatorial graphs has attracted considerable attention. The interest is stimulated in part by multiple existing and potential applications of analysis on graphs to information theory, signal analysis, image processing, computer sciences, learning theory, and astronomy.

My thesis is devoted to sampling, interpolation, approximation, and multi-resolution on graphs. The results in the existing literature concern mainly with these theories on unweighted graphs. My main objective is to extend existing theories and obtain new results about sampling, interpolation, approximation, and multi-resolution on general combinatorial graphs such as directed, undirected and weighted.

ACKNOWLEDGEMENTS

I want to take this opportunity to express my gratitude to all those who have helped me and inspired me during my studies.

I would like to express my deepest gratitude to my advisor, Prof. Isaac Pesenson, for giving me the opportunity to work with him, for his excellent guidance, continuous support, motivation, and providing me with an excellent atmosphere for doing research. He has always been very patient and enthusiastic and I have found his advice and support invaluable. I greatly value the freedom that he has given me in shaping my research work. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to gratefully acknowledge my wonderful committee members: Prof. Gerardo Mendoza and Prof. Shiferaw Berhanu, for their encouragement and inspiration to achieve the goal of completing my thesis. I am indebted to both of them for their continuous encouragement and support especially during my difficult times.

My sincere thanks also goes to Prof. Edward Letzter, head of our department and Prof. Martin Lorenz, our graduate chair for providing me with all the necessary support including the financial assistance(Teaching Assistantship) during my studies.

Most importantly, none of this would have been possible without the love, patience, and support of my family. My sons Bisrat and Brook, has been a constant source of love and strength all these years. I thank my mother, Tewabech Yimam, for the support and love she provided me through my entire life. I would also like to thank my younger brother Dr. Amare Abebe (Amex), for his encouragement, unconditional love and support. He was always there whenever I needed any technical support during my research and writing my thesis. His endless patience, kindness, love and brilliance is greatly appreciated.

Finally, I would like to acknowledge my friend and and former officemate, Shimao Fan, for his MatLab support.

To my father,

Abebe Gidelew,

whom I wanted to show my gratitude for how he has shaped my life and continues to motivate me everyday, but unfortunate to see this day.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Harmonic analysis is nowadays one of the most important branches of mathematical analysis. There are many applications of harmonic analysis concepts and techniques ranging from the most classical Fourier transform which has been used for decades by physicists, astronomers, engineers, etc., to the relatively new applications of wavelets to signal processing [1, 2, 3, 4, 5, 6]. Many signal processing techniques are based on transform methods. It is well known that wavelet transforms, Laplace transforms, and Fourier transforms are among some of the most common types of transforms in use. Although wavelets are a relatively recent development in mathematics, they have proved to be extremely useful tools for signal processing. Much of the power of wavelet methods comes from their ability to simultaneously locate signal content in both time and frequency (time-frequency localization) [6].

Graphs and networks have been successfully used in a variety of fields such as machine learning [20, 21], data mining [106], image analysis [80], complex networks [107] and social sciences that are confronted with the analysis and modeling of high-dimensional datasets. Harmonic analysis tools originally developed for Euclidean spaces and regular lattices are now being extended to the general settings of graphs and networks in order to analyze geometric and topological structures, and data and signals measured on them [1, 7, 8, 13, 19, 22, 23, 24, 25, 26, 27, 33, 44, 45, 73, 74, 79, 87, 32, 90]. The

development of this new analysis on combinatorial graphs attracted considerable attention. Many interesting scientific problems involve analyzing and manipulating structured data such as data mining. Such data often consist of sampled real valued functions defined on some irregular domain sets such as graphs.

As many traditional and classical methods for signal processing are designed for data defined on regular Euclidean spaces, the development of methods that are able to accommodate complicated data domains is also an important problem. Consequently, wavelets on graphs [1], the concepts of signal processing on graphs [7], uncertainty principles on graphs [19, 73], graph Laplacians [9, 10, 11, 13, 15, 16], Fourier transform on graphs, diffusion on graphs and sampling and approximation theory on graphs [22, 23, 24, 25, 26, 27, 33] have been developed by mathematicians and computer scientists in recent years.

For example, motivated by their effectiveness and wide range of applications, several mathematicians extended the theory of wavelet transforms for signals defined on graphs [1, 2, 3, 4, 5]. The approach on describing wavelets on graphs uses only the connectivity information encoded in the edge of weighted graphs, and does not rely on any other attributes of the vertices (such as their positions as embedded in High dimensional Euclidean spaces). As a result, the transform can be defined and calculated for any domain where the underlying relations between data locations can be represented by a weighted graph. This is important as a weighted graph provides a flexible model for approximating the data domains of a large class of problems. Some data sets can naturally be modeled as scalar functions defined on the vertices of graphs. For instance, computer networks, transportation networks, or social networks can all be described by weighted graphs, with the vertices corresponding to individual computers, cities or people respectively. The graph wavelet transform could be useful for analyzing data defined on the vertices, where the data is expected to be influenced by the underlying topology of the graph.

A spectral graph uncertainty principle (an uncertainty principle for signals

defined on graphs) which resembles the classical Heisenberg's famous uncertainty principle in time-frequency (see section 2.5) was also recently discussed in [19]. Various forms of the uncertainty principle in other settings have also attracted the attention of researchers in the past few years. For example, an uncertainty principle on homogenous trees [73], on the n-dimensional motion groups, nilpotent Lie groups, and non-compact semi-simple Lie groups, and symmetric spaces (see the references in [73]) have been studied in the last few years.

Recently a new approach was developed to the following topics on combinatorial graphs: Poincaré and Plancherel-Polya-type inequalities, sampling theory of Paley-Wiener functions, Lagrangian splines, Lagrange interpolation and approximation [22, 23, 24, 25, 26, 27, 33]. The results were obtained for both finite and infinite graphs. However, most of the previous works contain the assumption that graphs are unweighted, undirected, degrees of vertices are uniformly bounded. Although this assumption seems to be innocent, it imposes a strict restriction on weighted and directed graphs. The main goals of my research is to extend these notions to weighted and directed graphs.

Spectral graph theory is a useful subject which provides a bridge between the classical signal processing and the evolving field of graph signal processing. It studies the relation between graph properties and the spectrum of matrices associated to a graph, unlike algebraic graph theory which studies graphs by using algebraic properties of associated matrices. For details about special graph theory see [9, 10, 11, 13, 15, 16, 56] and for details about algebraic graph theory see [14]. I will particularly focus on graph Laplacian and its spectrum as it is the main tool in my research.

The thesis is organized in a such a way that the first sections up to section 4.2 give an overview of some relevant results from the classical Fourier analysis and recent results of analysis on graphs and sections from 4.3 up to the end contain my contributions to this subject. My own theoretical and experimental results are contained in the last 5 chapters.

In chapter 2, we briefly discuss the classical harmonic analysis, with main

emphasis on the Fourier transform and its applications, the classical Paley-Wiener spaces, Shannon Sampling Theorem, Heisenberg's uncertainty principle, and frame theory.

In chapter 3, we give an exposition of basic notations and concepts of graphs. In particular, we discuss about graphs and the Laplacian operators associated with graphs. Three most useful matrices associated with graphs, namely, the adjacency matrix, combinatorial Laplacian and the normalized Laplacian for both weighted and unweighted graphs will be discussed. We will also review the spectral properties of the Laplacian operators, known Laplacian eigenvalue bounds, spectral decomposition theorem, and the variational principle. Infinite graphs and an essential self-adjointness of an infinite Laplacian are also discussed here. In the case of infinite graphs, it is shown that the Laplacian operator is not always bounded and hence its analysis is more complicated. However, the Laplacian of a locally-finite (infinite) graph with an appropriate domain is a positive *essentially self-adjoint* operator.

Directed graphs and their Laplacian matrices are also discussed in this chapter. The Laplacian operator (matrix) on a directed graph is defined in terms of the transition probability matrix associated to the graph. A comparison of the directed and undirected Laplacians on regular graphs is also given.

In chapter 4, a graph Fourier transform which resembles in many ways to the classical Fourier transform is defined using the graph Laplacian eigenfunctions. In this case the Laplacian eigenvalues will be considered as the frequency and we will give an explanation and experimental results for that. We will then discuss a sampling theory on combinatorial graphs developed in [22] and further extend the idea to weighted and directed graphs. A graph setting of Paley-Wiener spaces and the sampling theorem will be discussed. It is shown that functions from some of these spaces are uniquely determined by their values on some subsets of the vertex set called uniqueness (sampling) sets. Uniqueness sets are described in terms of Poincaré-type inequalities. It is shown that every finite subset of a graph admits a Poincaré inequality and

is a uniqueness set for functions in some Paley-Wiener space.

In chapter 5, we will introduce a new method of signal approximation using frames. Frames are defined in terms of the Laplacian eigenfunctions. A multiresolution analysis on weighted graphs will be discussed. It is shown further in this chapter that every finite subset of the vertex set is a sampling set for some space of functions. Thus, the problem of signal interpolation can be posed as the problem of first defining the set of vertices S with known sample values as a uniqueness set for some $PW_\omega(G)$, and then reconstructing the signal values on the the complement set, $V \backslash S$. Given a subset S of the vertex set V, an algorithm to effectively compute the optimal frequency ω such that S is a uniqueness set for $PW_\omega(G)$ is developed. Many experimental results will be given in this chapter.

Then, in chapter 6, we will extend the idea of pointwise sampling to average sampling on a more general weighted graphs. The average value of a signal will be defined as a weighted average and an approximation technique will be introduced using average splines interpolating the signal on some disjoint subsets of the vertex set. Average splines are defined as minimizers of Sobolev norms which are introduced in terms of a combinatorial Laplace operator. It is shown that such splines interpolate functions on some subsets of the graph called *sampling sets* and provide *optimal* approximations to them.

Chapter 7 is devoted to establishing quadratures on combinatorial graphs. We develop a set of rules (quadratures) which allow for approximation or exact evaluation of "integrals" $\sum_{v \in V} f(v)$ of functions by using their values on subsets $U \subset V$ of vertices. Two types of quadratures are developed. Quadratures of the first type are exact on spaces of variational splines on graphs. Since bandlimited functions can be obtained as limits of variational splines we obtain quadratures which are essentially exact on spaces of bandlimited functions. Quadratures of the second type are exact on spaces of bandlimited functions. Accuracy of quadratures is given in terms of smoothness which is measured by means of combinatorial Laplace operator. The results have potential applications to problems that arise in data mining.

Finally, in chapter 8, we will further extend the application of splines in spectral graph drawing. More specifically, we will use splines approximating the lowest few eigenfunctions of the Laplacian to draw graphs and will explain the idea why these splines will be important in graph drawing. We will also discuss the high dimensional embedding and *eigen-projection methods*. We then construct a new low dimensional subspace of $\mathbb{R}^n$ for the coordinate axes to project graphs into low dimensions.

CHAPTER 2

CLASSICAL HARMONIC ANALYSIS

The Nyquist-Shannon sampling theorem, after Harry Nyquist and Claude Shannon, in the literature more commonly referred to as *Shannon's Sampling Theorem* or simply as the *sampling theorem*, is a fundamental result in the field of information theory, in particular telecommunications and signal processing. *Sampling* is the process of converting a signal (for example, a function of continuous time or space) into a numeric sequence (a function of discrete time or space).

It has been about six decades since Shannon introduced the sampling theorem to the communication theory. In 1949, Shannon published a paper titled Communication in the Presence of Noise, which set the foundation of the classical information theory [82]. This paper is considered one of the theoretical works that have the greatest impact on modern electrical engineering [83]. In order to formulate his rate/distortion theory, Shannon needed a general mechanism for converting an analog signal into a sequence of numbers. This led him to state the classical sampling theorem at the very beginning of his paper as follows.

Theorem 2.1 [*Shannon*] *If a function $f(x)$ contains no frequencies higher than B, it is completely determined by giving its ordinates at a series of points*

spaced $\frac{1}{2B}$ *seconds apart.*

A signal f is called *bandlimited* if it involves only frequencies smaller than some constant, say ω, that is, if its Fourier transform $\hat{f}$ vanishes outside the finite interval $[-\omega, \omega]$. In other words, $\hat{f}$ is compactly supported in $[-\omega, \omega]$. The signal f is then called ω-bandlimited. Thus, Shannon's sampling theorem states that a bandlimited function can be perfectly reconstructed from a countable sequence of samples if the bandlimit, ω, is not greater than half the sampling rate (samples per second). The theorem also leads to a formula for reconstruction of the original function from its samples. When the bandlimit is too high (or there is no bandlimit), the reconstruction exhibits imperfections known as *aliasing*. Of course, in practice, infinite sequences, perfect sampling, and perfect interpolation are all replaced by approximations, deviating from the ideal mathematical reconstruction.

The principal impact of the Shannon sampling theorem on information theory is that it allows the replacement of a continuous bandlimited signal by discrete sequence of its samples without the loss of any information. Also it specifies the lowest rate (also known as the Nyquist rate) of such sample values that is necessary to reproduce the original continuous signal. The proof of Shannon's sampling theorem depends on the Fourier transform of the signal, which we will discuss in the next section.

2.1 Fourier Transform and its applications

Fourier transform (FT) is named in honor of Joseph Fourier (1768-1830), one of the greatest names in the history of mathematics and physics. Mathematically speaking, the Fourier transform is a linear operator that maps a functional space to another functionals space and decomposes a function into another function of its frequency components [73]. In 1807, Fourier showed that any periodic function could be represented by a series of sinusoidal functions. The formulae used to define Fourier transform vary according to dif-

ferent authors. However, they are essentially the same but using different scales.

If f is an integrable function on $\mathbb{R}$, its Fourier transform is the function $\hat{f}$ on $\mathbb{R}$ defined by

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i\xi x}\, dx, \quad \xi \in \mathbb{R}. \tag{2.1}$$

It is also customary to write

$$\mathfrak{F}(f)(\xi) = \hat{f}(\xi)$$

for the Fourier transform of f.

Since $e^{-i\xi t}$ has absolute value 1, the integral converges absolutely for all ξ and defines a bounded function of ξ:

$$|f(\xi)| \leq \int_{-\infty}^{\infty} |f(x)|\, dx = ||f||_{L_1}.$$

Moreover, since $|e^{-i\xi x} f(x) - e^{-i\eta x} f(x)| \leq 2|f(x)|$, the dominated convergence theorem implies that $\hat{f}(\xi) - \hat{f}(\eta) \to 0$ as $\xi \to \eta$, that is $\hat{f}$ is continuous. In addition, If f is sufficiently smooth, then it can be reconstructed from its Fourier transform using the inverse Fourier transform

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i\xi x}\, d\xi. \tag{2.2}$$

Thus, the inverse Fourier transform is the procedure for recovering f from $\hat{f}$.

Strictly speaking, eq.(2.1) and eq.(2.2) are well defined only if f and $\mathfrak{F}(f)$, are absolutely integrable; for general L_2-functions f, $\mathfrak{F}(f)$ is defined via a limiting process.

Fourier transform, which was first proposed to solve PDEs such as Laplace, Heat and Wave equations, has enormous applications in physics, mathematics, engineering and chemistry. For example, some of its applications include:

1. *Signal analysis*: Fourier transform is essential to understand how a signal behaves when it passes through filters, amplifiers and communication

channels.

If $f(t)$ represents the amplitude of a signal such as a sound wave or an electromagnetic wave at time t, then the Fourier representation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i\xi t} \, d\xi$$

exhibits f as a continuous superposition of the simple periodic waves $e^{i\xi t}$ as ξ ranges over all possible frequencies.

2. *Image processing*: Transformation, representation, and encoding, smoothing and sharpening images. Image segmentation is one of the most widely studied problem in image analysis.

3. *Data analysis*: Fourier transform can be used as a high-pass, low-pass, and band-pass filters and it can also be applied to signal and noise estimation by encoding the time series. A *high-pass filter* is an electronic filter that passes high-frequency but attenuates (reduces the amplitude) signals with frequency lower than the cutoff frequency. A *low-pass filter* is a filter that passes low frequency signals and attenuates signals with frequencies higher than the cutoff frequency. A low-pass filter is the opposite of a high-pass filter. A *band-pass filter* is a combination of low-pass and high-pass filters.

As mentioned before, the proof of Shannon's sampling theorem depends on the Fourier transform of the signal. The reconstruction formula that complements Shannon's sampling theorem is

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\omega}\right) \frac{\sin(\omega t - n\pi)}{\omega t - n\pi}. \tag{2.3}$$

Shannon's proof starts by letting

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i\xi t} \, d\xi = \frac{1}{2\pi} \int_{-\omega}^{\omega} \hat{f}(\xi) e^{i\xi t} \, d\xi, \tag{2.4}$$

since $\hat{f}(\xi)$ is assumed to be zero outside the band $(-\omega, \omega)$.

2.2 Why bandlimited functions?

It is natural to ask why bandlimited functions can be completely reconstructed from their uniform samples. Here, we briefly explain the reason why that is generally true.

Suppose f represents a signal that we can measure its values at some sequence of times $t_1 < t_2 < \dots$. How much information can we gain this way? Clearly, for an arbitrary function $f(t)$, knowing a discrete set of values $f(t_1)$, $f(t_2), \dots$ tells us essentially nothing about the values of f at other points and it is difficult or even impossible to reconstruct f from these discrete set of values in a stable way. However, if f is known to involve only certain (low) frequencies (i.e., f is bandlimited), a lot can be said about the signal. The set of ω-bandlimited functions form the *Paley-Wiener class* $PW_\omega(\mathbb{R})$. Such functions f are restrictions to the real line of entire functions $F(z)$ of exponential type ω. Since $e^{i\xi t}$ does not change much on any interval of length $\Delta t \ll \xi^{-1}$, $f(t + \Delta t)$ will not differ much from $f(t)$, when $\Delta t \ll \omega^{-1}$; hence we should pretty well know f once we know the values $f(t_j)$ at a discrete sequence $\{t_j\}$ of points with $t_{j+1} - t_j \approx \omega^{-1}$. Thus, the classical sampling theorem says that if f is band-limited with $\widehat{f}(\xi) = 0$ for $|\xi| \geq \omega$, then f is completely determined by its values at the points $t_n = \frac{n\pi}{\omega}, n = 0, \pm 1, \pm 2, \dots$. It is an interesting observation that such functions are completely determined by their values at the points $x = \frac{n\pi}{\omega}$, where $n \in \mathbb{Z}$. In particular, we have,

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\omega}\right) \frac{\sin(\omega t - n\pi)}{\omega t - n\pi},$$

where convergence is understood in the L_2-sense. It is important to note that the functions

$$s_n(t) = \frac{\sin(\omega t - n\pi)}{\omega t - n\pi} \qquad (n = 0, \pm 1, \pm 2, \dots)$$

form an orthogonal basis for the space of ω-bandlimited functions, $PW_\omega(\mathbb{R})$ and the sampling formula (2.3) is merely the expansion of f with respect to

this basis. From the practical point of view, this expansion has the disadvantage that it generally does not converge very rapidly, since the sinc function $\frac{\sin x}{x}$ decays slowly as $x \to \infty$. However, a more rapidly convergent expansion for a function f can be obtained by *oversampling*, that is, by replacing the sequence of points $\frac{n\pi}{\omega}$ at which f is sampled by a more closely spaced sequence $\frac{n\pi}{\lambda\omega}$, $(\lambda > 1)$.

We know from Heisenberg's uncertainty principle (see Heisenberg's Inequality in sec 2.6) that it is impossible for a signal to be both bandlimited and time limited; that is it is impossible for f and $\hat{f}$ both to vanish outside a finite interval unless f is identically zero. Indeed, if $f \in L_2$ and $\widehat{f}(\xi) = 0$ for $|\xi| > \omega$, then the integral

$$F(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\xi z} \widehat{f}(\xi)\, d\xi$$

makes sense for any complex number z; moreover, we can differentiate under the integral to see that $F(z)$ is analytic. Thus, f is the restriction to the real axis of the entire analytic function F, and in particular, f cannot vanish except at isolated points unless it vanishes identically. In exactly the same way, if $f \neq 0$ vanishes outside a finite interval then $\hat{f}$ has only isolated zeros. These facts are aspects of the well known Heisenberg's uncertainty principle, which states that f and $\hat{f}$ cannot both be highly localized. That is, if f vanishes (or is very small) outside some small interval, then $\hat{f}$ has to be quite "spread out," and vice-versa.

2.3 Sampling non-bandlimited signals

The sampling theorem plays a crucial role in signal processing and communications: it tells us how to convert an analog signal into a sequence of numbers, which can then be processed digitally or coded on a computer. While Shannon's result is very elegant and has proven to be extremely useful, there are several problems associated with it [83]. First it is an idealization: real-

world signals or images are never exactly bandlimited. Second, there is no such device as an ideal (anti-aliasing or reconstruction) low-pass filter. Third, Shannon's reconstruction formula is rarely used in practice (especially with images) because of the slow decay of the sinc function. Instead, practitioners typically rely on much simpler techniques such as linear interpolation. Despite these apparent mismatches with the physical world, researchers continue to show that a reconciliation is possible and that Shannon's sampling theory, in its modern and extended versions, can perfectly handle such non-ideal situations.

Although Shannon's Sampling theorem was stated primarily for bandlimited functions which are uniformly sampled, it was later extended to a more general random functions and irregular samples [84, 85]. The sampling technique discussed in Shannon's sampling theorem and its extensions is explicit sampling in the sense that a bandlimited function $f(t)$ is represented in terms of its samples $f(t_n)$ at preselected instants $\{t_n\}$ which are independent of $f(t)$. On the other hand, many researchers have tried to extend the idea of Shannon to reconstruct bandlimited signals from irregularly sampled data and non-bandlimited signals from regularly sampling values. In the former case, one is asking whether and how a bandlimited signal f can be completely reconstructed from its irregularly sampled values $f(x_i)$. This has many applications in signal and image processing, seismology, meteorology, medical imaging, etc. (see [88] and the references therein.)

In the latter case, the standard signal processing practice is to apply a low-pass filter prior to sampling in order to suppress aliasing. To understand the modern point of view, we need to consider the Hilbert space $L_2(\mathbb{R})$, which consists of all functions that are square integrable (in Lebesgue's sense). The corresponding L_2-norm is

$$||f|| = \left(\int_{-\infty}^{+\infty} |f(x)|^2 \, dx \right)^{1/2} = \sqrt{\langle f, f \rangle}. \tag{2.5}$$

This norm is induced by the L_2 inner product

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(x)\, g^*(x)\, dx,$$

where $g^*(x)$ is the conjugate transpose of $g(x)$, and in fact it is just $g(x)$ if g is real.

We now assume that the input function that we want to sample is in $L_2(\mathbb{R})$, a space considerably larger than the subspace of bandlimited functions, $PW_\omega(\mathbb{R})$. This means that we will need to make an approximation if we want to represent a non-bandlimited signal in the bandlimited space $PW_\omega(\mathbb{R})$.

2.4 Projections and Least-Squares approximation

Definition 2.1 *Let ϕ be an element of a normed linear space V with $||\phi|| = 1$. For any $f \in V$ the projection of f in the direction of ϕ is denoted by $proj_\phi f$ with the definition*

$$proj_\phi f = \langle f, \phi\rangle \phi. \tag{2.6}$$

Theorem 2.2 *f is a scalar multiple of ϕ if and only if*

$$proj_\phi f = f$$

.

Proof: If $f = \alpha\phi$ for some scalar α, then we have

$$\text{proj}_\phi f = \langle \alpha\phi, \phi\rangle \phi = \alpha\langle \phi, \phi\rangle \phi = \alpha||\phi||^2\phi = \alpha\phi = f. \qquad (\text{since } ||\phi|| = 1)$$

Conversely, if $proj_\phi f = f$, then by (2.6) we have $\langle f, \phi\rangle\phi = f$, i.e., f is a scalar multiple of ϕ. ■

The next theorem is obvious.

Theorem 2.3 *If $\psi = f - proj_\phi f$, then the inner product $\langle \psi, \phi\rangle = 0$, and ψ is orthogonal to ϕ.*

Definition 2.2 *Let V denote a normed linear space, and let Ω_n denote a subspace of V spanned by an orthonormal sequence $\phi_1, \phi_2, \ldots, \phi_n$ in V. The projection of $f \in V$ into the subspace Ω_n is defined by*

$$proj_{\Omega_n} f = \sum_{\ell=1}^{n} proj_{\phi_\ell} f = \sum_{\ell=1}^{n} \lambda_\ell \phi_\ell, \text{ where } \lambda_\ell = \langle f, \phi_\ell \rangle. \tag{2.7}$$

Theorem 2.4 *$proj_{\Omega_n} f = f$ if and only if $f \in \Omega_n$.*

Proof: If $\text{proj}_{\Omega_n} f = f$, then by (2.7) f is a linear combination of $\phi_1, \phi_2, \ldots, \phi_n$ from the orthonormal sequence which spans Ω_n, hence $f \in \Omega_n$. Conversely, if $f \in \Omega_n$, then f can be written as

$$f = \sum_{k=1}^{n} \alpha_k \phi_k,$$

where the α_k's are scalars, and the ϕ_k's are from the orthonormal sequence spanning Ω_n. To obtain $\text{proj}_{\Omega_n} f$ according to (2.7), we compute the coefficient λ_ℓ defined by the inner product $\langle f, \phi_\ell \rangle$, namely,

$$\lambda_\ell = \langle f, \phi_\ell \rangle = \left\langle \sum_{k=1}^{n} \alpha_k \phi_k, \phi_\ell \right\rangle = \sum_{k=1}^{n} \alpha_k \langle \phi_k, \phi_\ell \rangle = \alpha_\ell \langle \phi_\ell, \phi_\ell \rangle = \alpha_\ell.$$

Then we have

$$\text{proj}_{\Omega_n} f = \sum_{\ell=1}^{n} \lambda_\ell \phi_\ell = \sum_{\ell=1}^{n} \alpha_\ell \phi_\ell = f. \qquad \blacksquare$$

Theorem 2.4 shows that for every $f \in \Omega_n$, one can express it as a linear combination of the elements from the orthonormal basis of Ω_n conveniently, because such an expression is given by $\text{proj}_{\Omega_n} f$ which explicitly defines each coefficient to be the inner product of f and an element from the orthonormal sequence. Notice that this theorem also shows that if $f \notin \Omega_n$, then $f \neq \text{proj}_{\Omega_n} f$.

The next theorem shows that if $\psi = f - \text{proj}_{\Omega_n} f$, then ψ is orthogonal to Ω_n.

Theorem 2.5 *Let Ω_n be the subspace of V spanned by the orthonormal sequence $\phi_1, \phi_2, \ldots, \phi_n$. If $\psi = f - proj_{\Omega_n} f$, then ψ is orthogonal to every element $g_n \in \Omega_n$.*

Proof: By Definition 2.2, $\text{proj}_{\Omega_n} f = \sum_{k=1}^{n} \lambda_\ell \phi_k$ for $\lambda_k = \langle f, \phi_k \rangle$. To show that ψ is orthogonal to every $g_n \in \Omega_n$, it suffices to show that $\langle \psi, \phi_\ell \rangle = 0$ for each $\ell = 1, 2, \ldots, n$ since every element $g_n \in \Omega_n$ can be expressed as

$$g_n = \sum_{\ell=1}^{n} \alpha_\ell \phi_\ell$$

for some scalars α_ℓ. Now for $\ell = 1, 2, \ldots, n$, we have

$$\begin{aligned}
\langle \psi, \phi_\ell \rangle &= \left\langle f - \sum_{k=1}^{n} \lambda_k \phi_k, \phi_\ell \right\rangle \\
&= \langle f, \phi_\ell \rangle - \sum_{k=1}^{n} \lambda_k \langle \phi_k, \phi_\ell \rangle \\
&= \langle f, \phi_\ell \rangle - \lambda_\ell \langle \phi_\ell, \phi_\ell \rangle \\
&= \langle f, \phi_\ell \rangle - \lambda_\ell = 0.
\end{aligned}$$

The theorem is proved. ■

As we mentioned before, Theorem 2.4 shows that if an element $f \notin \Omega_n$, then $f \neq \text{proj}_{\Omega_n} f$. Therefore, in general, we can only approximate an arbitrary function by a finite Fourier series. It is shown in the next theorem that if Ω_n is a subspace of V, then for every $f \in V$ and every $g_n \in \Omega_n$, the difference $||f - g_n||$ is minimized when $g_n = \text{proj}_{\Omega_n} f$. In other words, the best least-squares approximation to f in the subspace Ω_n is given by $g_n = \text{proj}_{\Omega_n} f$.

Theorem 2.6 *If Ω_n is a subspace of the normed V, and Ω_n is spanned by the orthonormal sequence $\phi_1, \phi_2, \ldots, \phi_n$, then for ever $f \in V$, the element $g_n \in \Omega_n$ for which $||f - g_n||$ is a minimum is $g_n = proj_{\Omega_n} f$.*

Proof: Suppose $||f - g||^2$ is minimum by $g_n = \sum_{k=1}^{n} \alpha_k \phi_k \in \Omega_n$, where the coefficients α_k to be determined. We proceed to evaluate the inner product

defining $||f - g_n||^2$, and we obtain

$$\begin{aligned}
\langle f - g_n, f - g_n\rangle &= \langle f, f\rangle - \langle f, g_n\rangle - \langle g_n, f\rangle + \langle g_n, g_n\rangle \\
&= ||f||^2 - \sum_{k=1}^{n} \bar{\alpha_k}\langle f, \phi_k\rangle - \sum_{k=1}^{n} \alpha_k\langle \phi_k, f\rangle + \sum_{k=1}^{n} |\alpha_k|^2 \\
&= ||f||^2 - \sum_{k=1}^{n} \bar{\alpha_k}\langle f, \phi_k\rangle - \sum_{k=1}^{n} \alpha_k\overline{\langle f, \phi_k\rangle} + \sum_{k=1}^{n} |\alpha_k|^2 \\
&= ||f||^2 - \sum_{k=1}^{n} \bar{\alpha_k}\lambda_k - \sum_{k=1}^{n} \alpha_k\bar{\lambda_k} + \sum_{k=1}^{n} \alpha_k\bar{\alpha_k} \qquad \text{where} \lambda_k = \langle f, \phi_k\rangle \\
&= ||f||^2 + \sum_{k=1}^{n} (\bar{\lambda_k}\lambda_k - \bar{\alpha_k}\lambda_k - \alpha_k\bar{\lambda_k} + \bar{\alpha_k}\alpha_k) - \sum_{k=1}^{n} \bar{\lambda_k}\lambda_k \\
&= ||f||^2 + \sum_{k=1}^{n} (\bar{\lambda_k} - \bar{\alpha_k})(\lambda_k - \alpha_k) - \sum_{k=1}^{n} |\lambda_k|^2 \\
&= ||f||^2 + \sum_{k=1}^{n} |\lambda_k - \alpha_k|^2 - \sum_{k=1}^{n} |\lambda_k|^2.
\end{aligned}$$

To minimize the right side, we focus only on the term involving the unknown α_k's, i.e., the term $\sum_{k=1}^{n} |\lambda_k - \alpha_k|^2$. Since this term is non-negative, its minimum value is zero, which is reached when $\alpha_k = \lambda_k$ for each $k = 1, 2, \ldots, n$. Therefore, $||f - g_n||$ is minimized by

$$g_n = \sum_{k=1}^{n} \alpha_k\phi_k = \sum_{k=1}^{n} \lambda_k\phi_k = \sum_{k=1}^{n} \langle f, \phi_k\rangle\phi_k = \text{proj}_{\Omega_n} f.$$

Moreover, $||f - g_n||^2 = \langle f - g_n, f - g_n\rangle = ||f||^2 - \sum_{k=1}^{n} |\lambda_k|^2 = ||f||^2 - ||g_n||^2$, where $\lambda_k = \langle f, \phi_k\rangle$. ■

Now substituting Ω_n by $PW_\omega(\mathbb{R})$ and V by $L_2(\mathbb{R})$, this orthonormality property greatly simplifies the implementation of the approximation process by which a function $f \in L_2\mathbb{R})$ is projected onto $PW_\omega(\mathbb{R})$. Specifically, the orthogonal projection operator $P_\omega : L_2(\mathbb{R}) \to PW_\omega(\mathbb{R})$ can be written as

$$P_\omega f = \sum_{k\in\mathbb{N}} \langle f, \phi_k\rangle\phi_k,$$

where the inner product $c_k(f) = \langle f, \phi_k\rangle$ represents the signal contribution along the direction specified by ϕ_k and $\{\phi_k\}_k$, $k = 1, 2, \ldots$ is an orthonor-

mal basis for $PW_\omega(\mathbb{R})$. The projection theorem ensures that this projection operation yields the minimum-error approximation of f into $PW_\omega(\mathbb{R})$:

$$\widetilde{f} = P_\omega f = \arg \min_{g \in PW_\omega(\mathbb{R})} ||f - g||^2$$

The projection interpretation of the sampling process has one big advantage: it does not require the bandlimited hypothesis and is applicable for any function $f \in L_2(\mathbb{R})$.

2.5 Heisenberg's Uncertainty Principle

The classical uncertainty principle (in one-dimension), which is also known as the Heisenberg - Pauli- Weyl inequality states that :
If $f \in L_2(\mathbb{R})$ and $a, b \in \mathbb{R}$ are arbitrary, then

$$\left(\int_{-\infty}^{\infty} (x-a)^2 |f(x)|^2 \, dx\right)^{1/2} \left(\int_{-\infty}^{\infty} (\omega - b)^2 |\hat{f}(\omega)|^2 \, d\omega\right)^{1/2} \geq \frac{1}{4\pi} ||f||_2^2, \quad (2.8)$$

where $\hat{f}$ is the Fourier transform of f. Intuitively, the uncertainty principle is an inequality that involves both f and $\hat{f}$.

For $f \in L_2(\mathbb{R})$ with $||f||_2 = 1$ define the standard deviations

$$\Delta_f x = \min_{a \in \mathbb{R}} \left(\int_{-\infty}^{\infty} (x-a)^2 |f(x)|^2 \, dx\right)^{1/2},$$

and

$$\Delta_f \omega = \min_{b \in \mathbb{R}} \left(\int_{-\infty}^{\infty} (\omega - b)^2 |\hat{f}(\omega)|^2 \, d\omega\right)^{1/2}.$$

$\Delta_f x$ is a measure for the signal duration, i.e., it measures the size of the essential support of f. Similarly, $\Delta_f \omega$ measures the essential bandwidth of the signal centered around the average frequency $\bar{\omega} = \int \omega |\hat{f}(\omega)|^2 \, d\omega$. If $\Delta_f x$ is finite, then it is minimized at its expected (average) value $\bar{x} = \int x |f(x)|^2 \, dx$. This classical Heisenberg Uncertainty Principle may be phrased in the language of Fourier transforms by saying roughly that both f and $\hat{f}$ cannot be well localized unless $f = 0$.

Now inequality (2.8) can be written in the following standard form:

$$\text{if} \qquad ||f||_2 = 1, \quad \text{then} \quad \Delta_f x \,.\, \Delta_f \omega \geq \frac{1}{4\pi} \tag{2.9}$$

From Inequality (2.9), we observe that the uncertainty principle limits the degree to which a function can be simultaneously localized in time and frequency. In other words, if a signal is highly localized in time (meaning $\Delta_f x$ is small), it must widely spread in frequency (i.e., $\Delta_f \omega$ is large), and vice versa.

2.6 Frame Theory

The theory of frames is due to Duffin and Schaeffer [100], and it was developed to address problems in non-harmonic Fourier series. Prior to Duffin and Schaeffer, these problems were concerned with finding criteria on real sequence $\{t_n\}$ so that the closed linear span, $\overline{span}\{e_{t_n}\}$, of exponentials $e_{t_n}(x) = e^{2\pi i t_n x}$ would be equal to the space $L^2[-\Omega, \Omega]$ of finite energy signals defined on $[-\Omega, \Omega]$.

Definition 2.3 *A sequence of vectors $\{h_j : j \in \mathbf{J}\}$ in a (separable) Hilbert space $\mathcal{H}$ with inner product $\langle ., . \rangle_{\mathcal{H}}$ is a frame for $\mathcal{H}$ if there exist constants $0 < A \leq B$ such that*

$$A||f||^2 \leq \sum_j |\langle f, h_j \rangle|^2 \leq B||f||^2 \;\; \textit{for all } f \in \mathcal{H}. \tag{2.10}$$

Any two constants A and B satisfying (2.10) are called frame bounds. If A = B, then $\{h_j : j \in \mathbf{J}\}$ is called a *tight frame*. Frame bounds are not unique in general. The *optimal lower frame bound* is the supremum over all lower frame bounds, and the *optimal upper frame bound* is the infimum over all upper frame bounds. A frame is *exact* if it is no longer a frame whenever any one of its elements is removed from the sequence.

A frame for a vector space equipped with an inner product allows each vector in the space to be written as a linear combination of the elements in the frame by relaxing the orthogonality and uniqueness of the decomposition.

That is, orthogonality and linear independence between the frame elements is not required. Intuitively, we may think of a frame as a basis to which we have added some more elements. As such, frames are more flexible tools to work with and gives us more freedom to choose from.

Examples: An orthonormal basis is a tight frame with frame bounds A = B = 1. The union of any two orthonormal bases is a tight frame with frame bounds A = B = 2. The union of an orthonormal basis with k arbitrary unit vectors is a frame with bounds A = 1 and B = k+1.

To understand frames and reconstruction methods better, we study some important associated operators.

Definition 2.4 *The frame operator of the frame* $\{h_j : j \in \mathbf{J}\}$ *is the function* $S : \mathcal{H} \to \mathcal{H}$ *defined as*

$$Sf = \sum_j \langle f, h_j \rangle h_j \tag{2.11}$$

for all $f \in \mathcal{H}$.

The synthesis operator or the reconstruction operator D is defined for a finite sequence $c = (c_j)_{j \in \mathbf{J}}$ *by*

$$Dc = \sum_{j \in \mathbf{J}} c_j h_j \ \in \mathcal{H}.$$

Notice that the frame operator S and the synthesis operator D have the relation $S = DD^*$.

The optimal frame bounds A, B for a frame $\{h_j\}$ with frame operator S are given by $A = ||S^{-1}||^{-1}$, $B = ||S||$, where $||.||$ is the operator norm of S. In particular, if $\mathcal{H}$ is finite dimensional, the optimal lower frame bound is the smallest eigenvalue of the associated frame operator, and the optimal upper frame bound is the largest eigenvalue. Moreover, if $\mathcal{H}$ has dimension n and $\{h_j\}_{j=1}^m$ is a tight frame and $||h_j|| = 1$ for all j, then the frame bound is $A = m/n$.

The lower frame bound implies that frames are complete. Indeed, if $f \in \mathcal{H}$ and $\langle f, h_j \rangle = 0$ for all j, then

$$A||f||^2 \leq \sum_j |\langle f, h_j \rangle|^2 = 0, \text{ so } ||f|| = 0, \text{ which implies } f = 0.$$

Moreover, in the case of finite diminutional Hilbert space $\mathcal{H}$, the removal of a vector h_k from a frame $\{h_j\}$ for $\mathcal{H}$ leaves either a frame or an incomplete set. More precisely, given a frame $\{h_j\}_{j=1}^m$ and a frame operators S, the following holds true:

(a) If $\langle h_k, S^{-1}h_k \rangle \neq 1$, then $\{h_j\}_{j \neq k}$ is still a frame for $\mathcal{H}$.

(b) If $\langle h_k, S^{-1}h_k \rangle = 1$, then $\{h_j\}_{j \neq k}$ is incomplete, where S is the associated frame operator.

Frames are not only complete, but they also provide a description of the whole Hilbert space $\mathcal{H}$. The upper bound estimate in eq.(2.10) is not usually a problem and it is not difficult to derive a reasonable estimate for B. However, the lower bound estimate is usually the difficult part. For in stance, we have the following upper frame bound estimate in the case of finite dimensional Hilbert space $\mathcal{H}$.

Let $\{h_j\}_{j=1}^m$ be a frame for a finite dimensional Hilbert space $\mathcal{H}$. We can use the Cauchy-Schwarz inequality to show that

$$\sum_{j=1}^m |\langle f, h_j \rangle|^2 \leq \sum_{j=1}^m ||h_j||^2 \, ||f||^2 = \left(\sum_{j=1}^m ||h_j||^2 \right) ||f||^2 \quad \text{for all } f \in \mathcal{H}.$$

So the upper frame bound condition is automatically satisfied by $B = \sum_{j=1}^m ||h_j||^2$. However, one can find a better upper frame bound than $\sum_{j=1}^m ||h_j||^2$. The following theorem gives a necessary and sufficient condition for a set of vectors to be a frame in the case of finite dimensional Hilbert spaces.

Theorem 2.7 *[65] Let $\mathcal{H}$ be a finite dimensional Hilbert space. A set of vectors $\{h_j\}_{j=1}^m$ in $\mathcal{H}$ is a frame for $\mathcal{H}$ if and only if span$\{h_j\}_{j=1}^m = \mathcal{H}$.*

We should note here that this theorem is not true if $\mathcal{H}$ is infinite dimensional.

Remark 2.1 *: Given operators $S, T : \mathcal{H} \to \mathcal{H}$ we write $S \preceq T$ if $\langle Sf, f\rangle \leq \langle Tf, f\rangle$ for all $f \in \mathcal{H}$.*

Theorem 2.8 *[65] Given a sequence $\{h_j\}$ in a Hilbert space $\mathcal{H}$, the following two statements are equivalent:*

(a) $\{h_j\}$ is a frame with frame bounds A, B.

(b) The frame operators $S : \mathcal{H} \to \mathcal{H}$ defined by $Sf = \sum_j \langle f, h_j\rangle h_j$ is a bounded linear operator with $AI \preceq S \preceq BI$.

Theorem 2.9 *[89] Suppose $\{h_j : j \in \mathbf{J}\}$ is a frame for $\mathcal{H}$ with frame operator S and frame bounds A, B. Then S maps $\mathcal{H}$ onto $\mathcal{H}$ and is a positive invertible, self-adjoint operator satisfying $AI \preceq S \preceq BI$ and $B^{-}I \preceq S^{-1} \preceq A^{-1}I$.*

To understand the convergence properties of the non-orthogonal series $\sum_j c_j h_j$ better, we exploit Theorem 2.9 further.

Corollary 2.1 *Let $\{h_j : j \in \mathbf{J}\}$ be a frame for $\mathcal{H}$. If $f = \sum_{j\in\mathbf{J}} c_j h_j$ for some $c \in \ell^2(\mathbf{J})$, then for every $\epsilon > 0$ there exists a finite subset $F_0 = F_0(\epsilon) \subseteq \mathbf{J}$ such that*

$$\|f - \sum_{j\in F} c_j h_j\| < \epsilon \quad \textit{for all finite subsets } F \supseteq F_0.$$

We say that the series $\sum_{j\in F} c_j h_j$ converges unconditionally to $f \in \mathcal{H}$.

Proof. Choose $F_0 \subseteq \mathbf{J}$ such that $\sum_{n\notin F} |c_j|^2 < \frac{\epsilon}{\sqrt{B}}$ for $F \supseteq F_0$. Let $c_F = c.\chi_F \in \ell^2(\mathbf{J})$ be the finite sequence with terms $c_{F,j} = c_j$ if $j \in F$ and $c_{F,j} = 0$ if $j \notin F$. Then $\sum_{j\in F} c_j h_j = Dc_F$ and

$$\begin{aligned}\|f - \sum_{j\in F} c_j h_j\| &= \|Dc - Dc_F\| \\ &= \|D(c - c_F)\| \\ &= \sqrt{B}\|c - c_F\| < \epsilon.\end{aligned}$$

As another consequence of Theorem 2.9 we obtain the following reconstruction formula for f from the frame coefficients $\langle f, h_j \rangle$.

Corollary 2.2 *If $\{h_j : j \in \mathbf{J}\}$ is a frame with frame bounds $A, B > 0$, then $\{S^{-1}h_j : j \in \mathbf{J}\}$ is a frame with frame bounds $B^{-1}, A^{-1} > 0$, the so-called dual frame. Moreover, every $f \in \mathcal{H}$ has a non-orthonormal expansion*

$$f = \sum_{j \in \mathbf{J}} \langle f, S^{-1}h_j \rangle h_j,$$

and

$$f = \sum_{j \in \mathbf{J}} \langle f, h_j \rangle S^{-1}h_j,$$

where both sums converge unconditionally in $\mathcal{H}$.

The series expansions in Corollary 2.2 are useful if it is possible to calculate the dual frame explicitly. Often it is more convenient and more efficient to employ an iterative reconstruction method, which is usually called the frame algorithm.

Frame Algorithm: Given a relaxation parameter $0 < \lambda < \frac{2}{B}$, set $\delta = \max\{|1 - \lambda A|, |1 - \lambda B|\} < 1$. Let $f_0 = 0$ and define recursively

$$f_{n+1} = f_n + \lambda S(f - f_n). \tag{2.12}$$

Then $\lim_{n \to \infty} f_n = f$ with a geometric rate of convergence, that is,

$$||f - f_n|| \leq \delta^n ||f||. \tag{2.13}$$

Observe that $f_1 = \lambda S f = \lambda \sum_j \langle f, h_j \rangle h_j$ contains the frame coefficients as input. This suffices to compute the further approximations f_n and to reconstruct f completely.

Proof. Since $AI \preceq S \preceq BI$, we obtain

$$(1 - \lambda B)I \preceq I - \lambda S \preceq (1 - \lambda A)I.$$

Therefore

$$||I - \lambda S|| \leq \max\{|1 - \lambda A|, |1 - \lambda B|\} = \delta < 1, \tag{2.14}$$

because $\lambda < \frac{2}{B}$. Assume that the error estimate (2.13) is true for $k = 1, \ldots, n$ (there is nothing to solve for n = 0). Then

$$\begin{aligned} ||f - f_{n+1}|| &= ||f - f_n - \lambda S(f - f_n)|| \\ &= ||(I - \lambda S)(f - f_n)|| \\ &\leq ||I - \lambda S||\ ||f - f_n|| \\ &\leq \delta\, \delta^n ||f||; \end{aligned}$$

so we are done. ■

CHAPTER 3

LAPLACIANS ON GRAPHS

3.1 Undirected and unweighted graphs

Graphs are extremely useful structures to describe and understand numerous problems. An undirected graph G consists of a vertex set $V = V(G)$ and an edge set $E = E(G)$, where an edge is an unordered pair of distinct vertices of G. We will usually use uv to denote an edge connecting the vertices u and v and view the graph G as a pair (V, E). If uv is an edge, then we say that u and v are *adjacent* or that v is a *neighbour* of u, and denote this by writing $u \sim v$. A vertex is *incident* with an edge if it is one of the two vertices of the edge. Two graphs G and H are equal if and only if they have the same vertex set and the same edge set. Two graphs G and H are *isomorphic* if there is a bijection, φ say, from $V(G)$ to $V(H)$ such that $u \sim v$ in G if and only if $\varphi(u) \sim \varphi(v)$ in H. We say φ is an isomorphism from G to H. It is normally appropriate to treat isomorphic graphs as if they were equal. It is important to note that in a graph, the positions of the vertices (also called nodes) and edges do not really matter - the only information it conveys is which pairs of vertices are joined by an edge. A graph is called *complete* if every pair of vertices are adjacent, and the complete graph on n vertices is denoted by K_n. The *degree* of a vertex v in an unweighted graph is the number of edges incident to v, and denoted by $d(v)$. For a finite graph, the maximum of all

$d(v),\ v \in V$ is called the degree of the graph and denoted by $\Delta(G)$. In the case of infinite graphs, $\Delta(G) = \sup\{d(v) :\ v \in V(G)\}$, and may not be finite. We say that a graph is *regular* if every vertex has the same degree, and *k-regular* if that degree is k. A graph is *simple* if there is no loop and there is at most one edge between two distinct vertices. Graphs that have multiple-edges or self-loops are often called *multi-graphs*. If G is a connected graph, then the distance between two vertices u and v is defined to be the length of the shortest path joining them in G. We denote this distance by $d(u, v)$. On any connected undirected graph, the graph distance is a metric, so that (V, d) is a metric space. In fact, the only thing that needs to be checked is the triangle inequality, $d(u, v) \leq d(u, x) + d(x, v)$ for all vertices u, x, v. But this follows from the fact that, if we want to go from u to v, then one possibility is to go via vertex x. The diameter of G denoted by $diam(G)$, is the maximum distance over all pairs of vertices in G.

A subgraph of a graph G is a graph H such that

$$V(H) \subseteq V(G),\ E(H) \subseteq E(G).$$

If $V(H) = V(G)$, we call H a *spanning subgraph* of G. Any spanning subgraph of G can be obtained by deleting some of the edges from G. A subgraph H of G is an *induced subgraph* if two vertices of $V(H)$ are adjacent in H if and only if they are adjacent in G. Any induced subgraph of G can be obtained by deleting some of the vertices from G, along with any edges that contain a deleted vertex. Thus an induced subgraph is determined by its vertex set: we refer to it as the subgraph of G induced by its vertex set. The number of induced subgraphs of G is equal to the number of subsets of $V(G)$. A graph with no edges (but at least one vertex) is called *empty*. A set of vertices that induces an empty subgraph is called an *independent set*. A path of length r from u to v in a graph is a sequence of $r+1$ distinct vertices starting with u and ending with v such that consecutive vertices are adjacent. If there is a path between any two vertices of a graph G, then G is *connected*. Alternatively, G is disconnected if we can partition its vertices into two non-empty sets, X

and Y say, such that no vertex in X is adjacent to a vertex in Y. In this case we say G is a disjoint union of the two subgraphs induced by X and Y. An induced subgraph of G that is maximal, subject to being connected, is called a *(connected) component* of G.

A *cycle* is a connected graph where every vertex has exactly two neighbors; the smallest cycle is the complete graph K_3. An *acyclic graph* is a graph with no cycles. A connected acyclic graph is called a *tree*, and an acyclic graph is called a *forest*. A spanning subgraph with no cycles is called a *spanning tree*. It is easy to see that a graph has a spanning tree if and only if it is connected. A graph $G = (V, E)$ is called *k-partite* if V admits a partition into k classes such that vertices in the same partition class must not be adjacent. Instead of 2-partite one usually says *bipartite*.

3.2 Undirected weighted graphs

A *weighted* graph is a graph in which a weight (typically a real number) has been assigned to every edge. In other words, a graph G is weighted when there is a function $w : E(G) \to \mathbb{R}^+$ which associates a positive value $w(u, v)$ with each edge $uv \in E(G)$. We denote a weighted graph by a triple (V, E, w). The function w is called a $weightfunction$ on G. Although it is possible to assign complex values for weights of edges, we consider only real and non-negative weights in this work. So we may associate to G a *weight matrix* (also called a *weighted adjacency matrix*), $W = (w_{vu})$, where $w_{vu} = w(v, u)$, weight of edge vu. For finite graphs with $|V(G)| = n$, W is an $n \times n$ matrix which has entries w_{vu} and satisfies the following:

1. $w_{vu} = w_{uv}$, *i.e.*, W is symmetric.

2. $w_{vu} \geq 0$ and $w_{vu} > 0$ if and only if v is adjacent to u in G. In particular, $w_{vv} = 0$ for each vertex v in V.

The (weighted) degree of a vertex $v \in V$ is defined by

$$\mu(v) = \sum_{u \in V} w(u, v).$$

If the graph is unweighted or every edge has a constant weight c, then $\mu(v)$ is simply the constant c times the number of vertices incident to the vertex v. Throughout the the thesis, we assume that G is connected, otherwise, the problem can be dealt independently for each connected component. Also note that, unweighted graphs are special cases of weighted graphs in which the weight of each edge is one and the (weighted) degree of each vertex is the number of edges incident to it.

Graphs are generic data representation forms which are useful for describing the geometric structure of data domains in numerous applications, including social, energy, transportation, sensor, and neuronal networks. The weight associated with each edge in the graph often represents the similarity (or the strength of the relationship) between the two vertices it connects. The connectivities and edge weights are either dictated by the physical nature of the problem at hand or inferred from the data. For instance, the edge weight may be inversely proportional to the physical distance between vertices in the network. The notion of weighted graphs is fundamental to many applications of graph theory. A number of natural interpretations of edge weight exist in various contexts. For example, in optimization, cost capacity is natural, in electrical network theory, edge may carry resistance or capacitance, in random walks each edge carries a probability of moving from one incident vertex to the other, and in geometric or analytical applications the use of edge length is often most appropriate.

When the edge weights are not naturally defined by an application, one common way to define the weight of an edge connecting vertices u and v is via a thresholded Gaussian kernel weighting function:

$$w(u, v) = \begin{cases} \exp\left(-\frac{[dist(u,v)]^2}{2\sigma^2}\right) & \text{if } \mathrm{dist}(u, v) \leq \kappa, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

for some parameters σ and κ. In eq.(3.1), $\text{dist}(u, v)$ may represent a graph distance $d(u, v)$, the shortest path connecting u and v, or the Euclidian distance between two feature vectors describing u and v if graph is embedded into Euclidean space. A second common method is to connect each vertex to its k-nearest neighbors based on the graph or Euclidean distances.

3.3 Undirected graph Laplacians

Given a graph G one can associate several matrices which record information about the graph (the vertices and how they are interconnected). A natural question then arises, given that we know the eigenvalues or eigenvectors of these matrices associated with the graph, what can we say about the graph? Spectral graph theory looks at answering questions of this type and the study of the relations between eigenvalues and structures of a graph is the heart of the subject. By studying the various spectra of matrices that can be associated with the graph it is possible to get information about the graph that might otherwise be difficult to obtain. For instance, a graph is connected if and only if the smallest eigenvalue ($\lambda_1 = 0$) of its Laplacian matrix has multiplicity one, a graph is bipartite if and only if the largest eigenvalue of the normalized Laplacian is equal to 2. If the graph is not connected, then the multiplicity of zero as an eigenvalue of the Laplacian matrix determines the number of connected components of the graph [9]. Graphs can be drawn by using few eigenvectors of some matrices associated with it [46]. Graph drawing will be discussed in chapter 8.

Among several matrices associated with graphs which capture information about the graphs, our primary focus will be on the three most common and important matrices, namely, the Adjacency Matrix, the Combinatorial Laplacian and the Normalized Laplacian.

Let $G = (V, E)$ be a simple undirected finite graph and consider a real-valued function f over V; $f : V \to \mathbb{R}$. This is simply a vector indexed by the vertices of G. Functions defined on graphs are often called *graph signals*. Let

$L_2(G)$ denote the space of all real valued functions $f : V \to \mathbb{R}$. The set of such functions forms a vector space over $\mathbb{R}$ of dimension n and hence is isomorphic to $\mathbb{R}^n$, where $n = |V(G)|$. So a function f in $L_2(G)$ can be represented as a vector in $\mathbb{R}^n$, where the i-th component of the vector represents the function value $f(i)$ at the i-th vertex in V. Thus, the space $L_2(G)$ can be endowed with the standard inner product in $\mathbb{R}^n$ as

$$\langle f, g \rangle_{L_2(G)} = \sum_{i \in V} f(i)g(i). \tag{3.2}$$

The corresponding norm in $L_2(G)$ is

$$||f|| = \sqrt{\langle f, f \rangle}.$$

The *adjacency matrix* of a graph G is the 0 -1 matrix $A = A(G)$ indexed by the vertex set V of G, where $A_{ij} = 1$ when there is an edge from i to j in G and 0 otherwise, i.e.,

$$A_{ij} = \begin{cases} 1 & \text{if } i \sim j, \\ 0 & \text{otherwise.} \end{cases}$$

The fundamental structure of a graph G can be captured by its adjacency matrix. The diagonal of A is always zero because no loops are allowed and A is symmetric because the graph is undirected. A power of the adjacency matrix has also a nice interpretation. If A is the adjacency matrix of a graph, the (i, j)-th entry of A^k is a nonnegative integer which is the number of paths of length k from node i to node j in the graph [72]. A generalization will be considered for weighted graphs where each edge $uv \in E(G)$ is associated with a positive weight $w(u, v)$ (see section 3.3.4). In this case, the nonzero entries of the adjacency matrix are replaced by the weights of the corresponding edges. While the adjacency matrix is the most natural matrix to associate with a graph, it is also believed that it is the least useful. Daniel Spielman [56], claims that eigenvalues and eigenfunctions are most useful when used to understand a natural operator or natural quadratic form and the adjacency matrix provides neither.

3.3.1 Unnormalized graph Laplacian

In this and the next two subsections we consider only unweighted graphs. Weighted graphs will be treated in section 3.3.4. Let D denote the diagonal matrix (also called the *degree matrix*) with the (i,i)-th entry D_{ii} having value $d(i)$, and let A be the adjacency matrix of G. The *unnormalized graph Laplacian(also know as graph Laplacian)*, L is a symmetric operator on $L_2(G)$ and defined [9] by

$$L = D - A. \tag{3.3}$$

Hence, the graph Laplacian operator $L : L_2(G) \to L_2(G)$ is a linear operator in the vector space $L_2(G)$. Moreover, for any function $f \in L_2(G)$ it satisfies

$$(Lf)(v) = \sum_{\substack{u \\ u \sim v}} (f(v) - f(u)) = d(v)f(v) - \sum_{\substack{u \\ u \sim v}} f(u), \tag{3.4}$$

for any $v \in V$, where the sum over $u \sim v$ indicates summation over all vertices u that are connected to the vertex v. The graph Laplacian operator can also be given explicitly by the matrix

$$L_{ij} = \begin{cases} d(i) & \text{if } i = j, \\ -1 & \text{if } i \sim j, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the diagonal elements of the Laplacian are the degree of the vertices and are strictly greater than 0 for connected graphs. For any $f \in L_2(G)$, it can be verified that the quadratic form of the Laplacian is

$$\langle f, Lf \rangle = \sum_{i \sim j} (f(i) - f(j))^2. \tag{3.5}$$

The quadratic form measures the *smoothness* of the function f. It will be small if the function f does not jump too much over any edge [21, 56]. Applying the variational principle on self-adjoint operators, eq.(3.5) also immediately shows that the Laplacian is a nonnegative operator, and thus all eigenvalues are non-negative.

The Laplacian matrix of a graph and its eigenpairs can be used in several areas of mathematical research and have a physical interpretation in various physical and chemical theories. For example, eigenvectors of the Laplacian matrix can be used in clustering [41, 42, 43], semi-supervised learning [20, 21], non-linear dimensionality reduction, ranking, cutting, graph drawing [46, 47, 48, 49, 50, 51, 52, 53], and more. Although the associated adjacency matrix of a graph and its eigenvalues were much more investigated and widely used in the past than the Laplacian matrix, it is believed that the Laplacian spectrum is much more natural and more important than the adjacency matrix spectrum [15, 56].

An Important fact from Linear Algebra: If M is an n-by-n symmetric matrix, then there exist n mutually orthogonal unit vectors $\psi_1, \psi_2, \ldots, \psi_n$ and numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that ψ_i is an eigenvector of M of eigenvalue λ_i, for each i. This is a well known fact about symmetric matrices and can be found in any standard linear algebra books (e.g. see [72]). Another important property of symmetric matrices is if λ and μ are eigenvalues of M, with $\lambda \neq \mu$, then the eigenvectors corresponding to λ and μ are orthogonal.

Since both D and A are symmetric it is clear that the Laplacian is symmetric and also by eq.(3.5), L is positive semi-definite. Thus, it has n eigenvalues and all eigenvalues are real and nonnegative with $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ (repeated according to their multiplicities). The eigenvectors of L are also called the *eigenfunctions*. Notice that $L\mathbf{1} = \mathbf{0}$ where $\mathbf{1} = (1, 1, \ldots, 1)$. So 0 is an eigenvalue of L corresponding to the eigenfunction $\mathbf{1}$. Moreover, the multiplicity of the value zero as an eigenvalue of L is equal to the number of connected components of G. In addition, according to the above discussion, if the graph is connected then every eigenfunction corresponding to the eigenvalue $\lambda_i,\ i \geq 2$ is orthogonal to $\mathbf{1}$. It is also known that if λ_2 is big, then G is very well connected. Because the graph Laplacian L is a real symmetric matrix, it has a complete set of orthonormal eigenvectors, which we denote by

$\{\phi_\ell\}_\ell,\ \ell = 1, 2, \ldots, n$ with the associated eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ satisfying

$$L\phi_\ell = \lambda_\ell \phi_\ell$$

for $\ell = 1, \ldots, n$. We denote the *spectrum* of L by $\sigma(L)$. The spectrum $\sigma(L) \subseteq [0, 2\Delta(G)]$ (see for example [13]).

By the Spectral Decomposition (see Theorem 3.4),
for every function $f : V \to \mathbb{R}$ we have

$$f = \sum_{j=1}^{n} \langle f, \phi_j \rangle \phi_j.$$

Thus,

$$Lf = \sum_{j=1}^{n} \lambda_j \langle f, \phi_j \rangle \phi_j. \tag{3.6}$$

3.3.2 Normalized graph Laplacian

When studying spectral graph theory, it is sometimes useful to normalize the graph Laplacian by its degrees. The *normalized Laplacian* of G is defined by

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2},$$

with the convention that $D^{-1}(v, v) = 0$ for $d(v) = 0$. In some cases it is easier and more convenient to work with such normalized form. As in the case of the unnormalized Laplacian, the normalized Laplacian, $\mathcal{L}$ can be written explicitly (as a matrix) as:

$$\mathcal{L}_{uv} = \begin{cases} 1 & \text{if } u = v \text{ and } d(u) > 0 \\ -\frac{1}{\sqrt{d(u)d(v)}} & \text{if } u \sim v \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

If the graph G is connected, then for any $f \in L_2(G)$ and any $v \in V(G)$, the operator $\mathcal{L}$ satisfies

$$\mathcal{L}f(v) = \frac{1}{\sqrt{d(v)}} \sum_{\substack{u \\ u \sim v}} \left(\frac{f(v)}{\sqrt{d(v)}} - \frac{f(u)}{\sqrt{d(u)}} \right). \tag{3.8}$$

As in the case of the unnormalized Laplacian, zero is an eigenvalue of the normalized Laplacian operator. Moreover, if the vertices of the graph are labeled by $v_1, v_2, \ldots, v_n$, then the eigenvector for $\mathcal{L}$ corresponding to the eigenvalue 0 is $\psi_0 = (\sqrt{d(v_1)}, \sqrt{d(v_2)}, \ldots, \sqrt{d(v_n)})$. In this case, the spectrum of $\mathcal{L}$ is contained in $[0, 2]$ and 2 is an eigenvalue of $\mathcal{L}$ if and only if the graph is bipartite.

3.3.3 Weighted graph Laplacian

So far we have discussed about the combinatorial Laplacian and its normalized form on simple unweighted graphs. However, it is straightforward to generalize these concepts to weighted graphs. Let $G = (V, E, w)$ be a weighted graph and let W be its associated (weighted) adjacency matrix. Let D denote the diagonal matrix with the (v, v)-th entry having value $\mu(v) = \sum_{u \in V} w(v, u)$. The *weighted graph Laplacian* L_w is an operator on $L_2(G)$ defined [10] as

$$L_w = D - W.$$

The matrix form of L_w is given by the following explicit form

$$L_w(u, v) = \begin{cases} \mu(v) & \text{if } u = v \ , \\ -w(u, v) & \text{if } u \sim v, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

where $\mu(v) = \sum_{u \in V} w(u, v)$, the (weighted) degree of the vertex v. In particular, for any $f \in L_2(G)$, we have

$$(L_w f)(v) = \sum_{\substack{u \\ u \sim v}} w(v, u) \left(f(v) - f(u) \right), \quad f \in L_2(G). \tag{3.10}$$

Note that $L_w\mathbf{1} = 0$. Conversely, any symmetric matrix M with nonpositive off diagonal entries such that $M\mathbf{1} = 0$ is a weighted Laplacian.

Consider in $L_2(G)$ the following inner product: for any two functions $f, g \in L_2(G)$, set

$$\langle f, g\rangle := \sum_{v\in V} f(v)g(v)\mu(v).$$

The corresponding norm in $L_2(G)$ is

$$||f|| = \sqrt{\langle f, f\rangle}.$$

As in the case of unweighted Laplacian, L_w is symmetric and positive definite with respect to the above inner product.

For any $f \in L_2(G)$, the quadratic form of the weighted Laplacian is

$$\langle f, L_w f\rangle = \sum_{v\sim u} w(u, v)(f(v) - f(u))^2. \tag{3.11}$$

The quadratic form measures the smoothness of the function f. The weighted graph Laplacian arises in many applications (e.g see [10] and the references therein).

The following theorem (from [16]) lists some important facts and bound of the eigenvalues of L_w. The eigenvalues λ_1 and λ_n can be bounded in terms of the minimum and maximum degrees of G.

Theorem 3.1 *Let $G = (V, E, w)$ be a weighted graph of order n and L_w be its weighted Laplacian with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then*

(a) $\mathbf{1} = (1, 1, \ldots, 1)$ *is an eigenfunction of L_w with corresponding eigenvalue 0.*

(b) the multiplicity of 0 as an eigenvalue of L_w is equal to the number of connected components of G.

(c) $\lambda_2 \leq \frac{n}{n-1}\delta(G)$, *where* $\delta(G) = \min\{\mu(v);\, v \in V(G)\}$.

(d) $\frac{n}{n-1}\Delta(G) \leq \lambda_n \leq 2\Delta(G)$, *where* $\Delta(G) = \max\{\mu(v);\, v \in V(G)\}$. *If* G *is an unweighted graph then the last inequality of theorem can be strengthened to* $\lambda_n \leq \max\{d(u) + d(v);\, uv \in E(G)\}$. *If* G *is connected, then the equality holds if and only if* G *is bipartite.*

The normalized form of the weighted graph Laplacian of G is defined to be

$$\mathcal{L}_w = D^{-1/2} L_w D^{-1/2}.$$

In other words, we have

$$\mathcal{L}_w(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d(v) \neq 0, \\ -\frac{w(u,v)}{\sqrt{\mu(u)\mu(v)}} & \text{if } u \sim v, \\ 0 & \text{otherwise.} \end{cases}$$

Notation: Abusing the notation, for the sake of simplicity and uniformity, we will use the notation $\mathcal{L}$ for the normalized Laplacian for both weighted and unweighted graphs from now onwards and we will specifically mention it if there is a need to differentiate the two.

Theorem 3.2 *[16] For any finite, connected, weighted graph* $G = (V, E, w)$ *with* $|V| = n > 1$, *the following are true.*

(i) Zero is a simple eigenvalue of $\mathcal{L}$, *with its corresponding eigenfunction* $(\sqrt{\mu(v_1)}, \ldots, \sqrt{\mu(v_n)})$, *where* $v_1, \ldots, v_n$ *are vertices of the graph.*

(ii) All the eigenvalues of $\mathcal{L}$ *are contained in [0,2].*

(iii) If G *is not bipartite then all the eigenvalues of* $\mathcal{L}$ *are in [0,2).*

Theorem 3.3 *If* G *is a connected weighted graph with diameter* $d = d(G)$, *then* $\mathcal{L}$ *has at least* $d + 1$ *distinct eigenvalues.*

Proof: Let M be any nonnegative symmetric matrix with rows and columns indexed by V(G) and such that for distinct vertices u, v in V(G) we have $M_{uv} > 0$ if and only if $u \sim v$. Let the distinct eigenvalues of M be $\lambda_1, \lambda_2, \ldots \lambda_r$.

Then $M - \lambda_i I = \mathbf{0}$ for each $1 \leq i \leq r$, where $\mathbf{0}$ is the the zero vector. In particular, $(M - \lambda_i I)_{uv} = 0$ for any pair of vertices $u, v \in V(G)$ and $1 \leq i \leq r$. Hence $(M - \lambda_1 I)...(M - \lambda_r I) = 0$, which then implies M^r is a linear combination of $I, M, ..., M^{r-1}$. But if there exist $u, v \in V(G)$ such that $\mu(u, v) = r$, then $(M^i)_{uv} = 0$ for $0 \leq i \leq r-1$ and $(M^r)_{uv} > 0$, a contradiction. Hence $r > d$. Now let $M = nI - \mathcal{L}$, and the desired result follows. ■

3.3.4 Generalized Graph Laplacian

A symmetric matrix $M = M(G)$ is called a *generalized Laplacian* (or *discrete Schrödinger operator*) of G if it has nonpositive off-diagonal entries and for $u \neq v$, $M_{uv} < 0$ if and only if the vertices u and v are adjacent. On the other hand, for every symmetric matrix with nonpositive off-diagonal entries there exists a graph where two distinct vertices u and v are adjacent if and only if $M_{uv} < 0$. Similarly to eq.(3.4) and eq.(3.10) for any generalized Laplacian M and any $f \in L_2(G)$,

$$(Mf)(v) = \sum_{\substack{u \\ u \sim v}} (-M_{uv})[f(v) - f(u)] + p(v)f(v), \tag{3.12}$$

where

$$p(v) = M_{vv} + \sum_{u \sim v} M_{vu}$$

can be viewed as some potential on vertex v.

The quadratic form of the the generalized Laplacian can then be computed as

$$\langle f, Mf \rangle = \sum_{v \sim u} -M_{uv}(f(v) - f(u))^2 + \sum_{v \in V} p(v)(f(v))^2.$$

In fact, using eq.(3.12), for any $f \in L_2(G)$ we have

$$\begin{aligned}
\langle f, Mf\rangle = \sum_{v\in V} f(v)Mf(v) &= \sum_{v\in V} f(v)\left(\sum_{\substack{u\\ u\sim v}}(-M_{uv})[f(v)-f(u)] + p(v)f(v)\right)\\
&= \sum_{v\in V}\left(\sum_{\substack{u\\ u\sim v}}(-M_{uv})f(v)[f(v)-f(u)] + p(v)f^2(v)\right)\\
&= \sum_{v\in V}\sum_{\substack{u\\ u\sim v}}(-M_{uv})f(v)[f(v)-f(u)] + \sum_{v\in V} p(v)f^2(v)\\
&= \sum_{v,u\in V}(-M_{uv})f(v)[f(v)-f(u)] + \sum_{v\in V} p(v)f^2(v)\\
&= \sum_{u\sim v}(-M_{uv})(f(v)-f(u))^2 + \sum_{v\in V} p(v)f^2(v)
\end{aligned}$$

as desired. ■

Clearly, the graph Laplacians defined in eq.(3.4), (3.8), and (3.10) are special cases of eq.(3.12). For instance, the weighted graph Laplacian is a special case of eq.(3.12), with $-M_{uv} = w(u,v)$ and $p(v) = 0$ for all $v \in V$.

The following fundamental theorem follows from the spectral theory for symmetric operators.

Theorem 3.4 *[Spectral Decomposition] For a generalized graph Laplacian M on a finite graph G, there exists an orthonormal basis of $\mathbb{R}^n$ (which is isomorphic to $L_2(G)$) that consists of eigenfunctions $f_1, \ldots, f_n$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. Moreover, for every function $g : V \to \mathbb{R}$ we have*

$$Mg = \sum_{i=1}^{n} \lambda_i \langle g, f_i\rangle f_i,$$

and for the quadratic form,

$$\langle g, Mg\rangle = \sum_{i=1}^{n} \lambda_i \langle g, f_i\rangle^2.$$

Proof. Let $\{f_1, \ldots, f_m\}$ be an orthonormal (and hence linearly independent) set of $m < n$ eigenfunctions of M, and let E be the subspace that they span.

Since M has at least one eigenfunction (any symmetric matrix has at least one eigenvector), $m \geq 1$. The subspace E is M-invariant, and hence $E^{\perp}$ is M-invariant (given any real symmetric $n \times n$ matrix M, if U is a M-invariant subspace of $\mathbb{R}^n$, then $U^{\perp}$ is also M-invariant), and so $E^{\perp}$ contains a (normalized) eigenfunction f_{m+1}. Then $\{f_1, \ldots, f_m, f_{m+1}\}$ is an orthonormal set of $m+1$ eigenfunctions of M. Therefore, a simple induction argument shows that a set consisting of one normalized eigenfunction can be extended to an orthonormal basis consisting of eigenfunctions of M. ■

One of the consequences of this theorem is that the *algebraic multiplicity* of any eigenvalue λ, that is, its multiplicity as a root of the characteristic polynomial, coincides with the *geometric multiplicity*, that is the maximal number of linearly independent eigenfunctions with the same eigenvalue $\lambda (= dim\, ker(M - \lambda I))$. The following corollary is an immediate consequence of the theorem.

Corollary 3.1 *[The Variational Principle] Let $f_1, \ldots, f_n$ denote orthogonal eigenfunctions corresponding to the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ of a generalized graph Laplacian M. Let $F_k = \{f_1, \ldots, f_k\}$ be the set of the first k eigenfunctions and $F_k^{\perp}$ its orthogonal complement. Then*

$$\lambda_k = \min_{g \in F_{k-1}^{\perp}} \mathcal{R}(g) = \min_{g \in F_{k-1}^{\perp}} \frac{\langle g, Mg \rangle}{\langle g, g \rangle}, \tag{3.13}$$

where $\mathcal{R}(g) = \frac{\langle g, Mg \rangle}{\langle g, g \rangle}$ is the Rayleigh Quotient of g with respect to the general Laplacian M. Moreover, $\mathcal{R}(g) = \lambda_k$ for some $g \in F_{k-1}^{\perp}$ if and only if g is an eigenfunction corresponding to λ_k.

Proof. Every function $g \in F_{k-1}^{\perp}$ can be written as

$$g = \sum_{i=k}^{n} c_i f_i$$

for some c_i. Consequently,

$$\langle g, Mg \rangle = \sum_{i=k}^{n} \lambda_i c_i^2 \quad \text{and} \quad \langle g, g \rangle = \sum_{i=k}^{n} c_i^2.$$

Hence the the Rayleigh quotient of g satisfies the inequality

$$\mathcal{R}(g) = \frac{\langle g, Mg\rangle}{\langle g, g\rangle} = \frac{\sum_{i=k}^{n} \lambda_i c_i^2}{\sum_{i=k}^{n} c_i^2} \geq \frac{\sum_{i=k}^{n} \lambda_k c_i^2}{\sum_{i=k}^{n} c_i^2} = \lambda_k$$

and the equality holds if and only if all terms with eigenvalues $\lambda_i > \lambda_k$ vanish. Thus the result follows. ■

Corollary 3.2 *[Minimax-Theorem] Let $\mathcal{W}_k$ denote the set of subspaces of $\mathbb{R}^n$ of dimension at least k. Then*

$$\lambda_k = \min_{W \in \mathcal{W}_k} \max_{0 \neq g \in W} \frac{\langle g, Mg\rangle}{\langle g, g\rangle}. \tag{3.14}$$

Proof. Every function $g \in L_2(G)$ can be written as

$$g = \sum_{i=1}^{n} c_i f_i$$

for some c_i, where $\{f, \ldots, f_n\}$ is the orthonormal basis of eigenfunctions from theorem 3.4. Hence

$$\mathcal{R}(g) = \frac{\langle g, Lg\rangle}{\langle g, g\rangle} = \frac{\sum_{i=1}^{n} \lambda_i c_i^2}{\sum_{i=1}^{n} c_i^2}.$$

Then for every $W \in \mathcal{W}_k$, we can find some $g \in W$ where $c_1 = \cdots = c_{k-1} = 0$ and thus,

$$\sup_{g \in W} \mathcal{R}(g) \geq \sup_{g \in W,\, c_1 = \cdots = c_{k-1} = 0} \frac{\sum_{i=k}^{n} \lambda_i c_i^2}{\sum_{i=k}^{n} c_i^2} \geq \lambda_k.$$

Consequently,

$$\inf_{W \in \mathcal{W}_k} \sup_{g \in W} \mathcal{R}(g) \geq \lambda_k.$$

The equality holds if W is the subspace that is spanned by the first k eigenfunctions. This completes the proof. ■

Corollary 3.3 *[Diagonalization of Symmetric Operators] Given a Laplacian matrix L of a graph G, there are matrices Σ and Λ such that $\Sigma^T\Sigma = \Sigma\Sigma^T = I$ and $L = \Sigma^T\Lambda\Sigma$, where Λ is the diagonal matrix of eigenvalues of L(i.e., $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$) and Σ is the matrix whose columns are orthonormal basis of eigenfunctions of L.*

Remark 3.1 *The choice of unnormalized or normalized Laplacian makes no essential difference in our subsequent analysis although we may favor the latter because it leads to simpler expressions. We will mention it specifically whenever it is necessary.*

Laplacian matrices of graphs are closely related to the continuous Laplacian operator, the second order differential operator $\Delta(f)$ = -div(grad f). In fact, Agaskar and Lu [19] claimed that when the underlying graph is a line or a cycle, L provides the standard stencil approximation for the second order differential operator and the same holds for higher-dimensional lattice. In more general settings where the graphs are formed by sampling an underlying continuous manifold, the Laplacian matrix converges at high sampling densities to the Laplace-Beltrami operator. It approximates the Laplace-Beltrami operator on a compact manifold in the sense that if the dataset is large and samples uniformly random on a Low-dimensional manifold then the graph Laplacian acting on smooth functions on this manifold is a good discrete approximation that converges pointwise and uniformly to the elliptic Laplace-Beltrami operator applied to this function as the number of points goes to infinity [93]. This relationship yields an important bilateral link between the spectral geometry of Riemannian manifolds and graph theory and, makes it possible to use results about graphs in the study of Laplacians on manifolds and, conversely, to transfer results about Laplacians on manifolds to graphs.

3.4 Laplacians on infinite graphs

In the previous sections we considered only finite graphs. We now turn our attention to infinite graphs. Here, there are two very different areas of research: the first for locally-finite graphs in which each vertex has a finite degree, and the second for general graphs. Infinite graphs are considered in many domains in mathematics such as combinatorial and geometrical group theory, number theory, general and algebraical topology, set theory, probability

and mathematical physics. They are also considered in medical researches (brain cells, blood veins), chemistry, informatics, electrical networks, synthetic imagery, internet connections, telecom, social sciences and more (see [18, 74, 75] and the references therein)

Laplacians on graphs have been studied for a long time. However, much of the research has been devoted to finite graphs and bounded Laplacians. In contrast to the Laplacian on a finite graph, the Laplacian on an infinite graph is not always a bounded operator and hence, its analysis is more complicated. Certain properties related to unboundedness of the associated Laplacian on infinite graphs have become a focus of attention in recent years [18, 74, 75, 76, 77, 78, 79]. Further, it has been shown in [79] that in the case of locally-finite infinite graphs the Laplacian with an appropriate domain is a *positive essentially self-adjoint* operator. More precisely, the Laplacian is defined on the dense subspace in the Hilbert space $L_2(G)$ and hence its closure is self-adjoint in $L_2(G)$. This is an analogue of the well known result that the Laplacian initially defined on the set of smooth functions with compact support on a compact Riemannian manifold M extends to an unbounded self-adjoint operator on $L^2(M)$.

We recall that an unbounded symmetric linear operator on a Hilbert space is essentially self-adjoint if it has a unique self-adjoint extension. Keller [75] also pointed out that the spectral properties of infinite Laplacians might be very different from the well studied spectral properties of combinatorial Laplacians on finite graphs. Next, we will further investigate infinite Laplacians.

Let G = (V, E) be an undirected countably infinite graph, and let $\mathbf{M}$(G) = $\mathbf{M}$ = $(m_{uv})_{u,v\in V}$ be a square matrix indexed by the vertices of G. The spectrum of $\mathbf{M}$(G) depends on the choice of a suitable space on which $\mathbf{M}$ acts as a linear operator. Usually, one considers the Hilbert space of square summable functions

$$L_2(G) = \{f : V \to \mathbb{R} \mid \sum_{v\in V} |f(v)|^2 < \infty\}, \tag{3.15}$$

with inner product

$$\langle f, g\rangle = \sum_{v\in V} f(v)g(v),$$

and the associated norm

$$||f||_2 = ||f||_{L_2(G)} = \langle f, f\rangle^{1/2}.$$

For $u \in V$, denote by $\mathbf{e}_u$ the unit vector in $L_2(G)$ whose u-th entry is equal to one, all other entries are zero (i.e., $\mathbf{e}_u = \delta_{vu}$, the Kronecker delta supported at vertex u). Then $\{\mathbf{e}_u | u \in V\}$ is a complete orthonormal system for $L_2(G)$. For a weighted graph $G = (V, E, w)$,

$$L_2(G) = \{f : V \to \mathbb{R} \mid \sum_{v\in V} \mu^2(v)|f(v)|^2 < \infty\}, \tag{3.16}$$

where

$$\mu(v) = \sum_{\substack{u \\ u\sim v}} w(u, v).$$

The vector space $L_2(G)$ has a Hilbert structure, when endowed with the inner product

$$\langle f, g\rangle_{L_2(G)} = \sum_{v\in V} \mu(v)f(v)g(v). \tag{3.17}$$

$L_2(G)$ can also be replaced by any of the spaces $L_p(G)$, consisting of all real valued functions f on V satisfying

$$||f||_p = \left(\sum_{v\in V} |f(v)|^p\right)^{1/p} < \infty,$$

where $1 \leq p < \infty$; for $p = \infty$, the norm reduces to $||f||_\infty = \max\{|f(v)|,\ v \in V\}$. The action of $\mathbf{M}$ is matrix multiplication: the coordinates of $g = \mathbf{M}f$ are

$$g(u) = \sum_{v\in V} m_{uv}f(v),\ u \in V, \tag{3.18}$$

whenever the series converges. In the next section, our primary focus will be on locally-finite infinite graphs.

3.5 Locally-finite infinite graphs

We say that an infinite graph G is *locally-finite* if each vertex is of finite degree, i.e., $\mathrm{d}(u) < \infty$ for every vertex $u \in V$. The following are good examples of locally-finite graphs.

Example 1: Consider a weighted graph $G = (V, E, w)$ such that the cardinality of the set $\{u : w(v, u) \neq 0; v \in V\}$ is finite, i.e

$$|\{u : w(v, u) \neq 0\}| < \infty \quad \forall v \in V.$$

G is locally-finite.

Example 2: The infinite graph with vertices at the integers $\mathbb{Z}$ such that each vertex $n \in \mathbb{Z}$ is connected to vertices $n - 1$ and $n + 1$. This is a uniformly bounded locally finite (infinite) graph. This graph has been considered by many authors and its graph Laplacian is well studied.

Let G = (V, E) be a locally-finite countable graph with the (labeled) vertex set $V(G) = \{v_j : j \in \mathbb{N}\}$, where $\mathbb{N}$ is the set of natural numbers. As in the case of finite graphs, there are several ways to associate a matrix $\mathbf{M}$(G) with G. An obvious and most natural choice is the (infinite) adjacency matrix $\mathbf{A} = \mathbf{A}(\mathrm{G}) = (a_{ij})$. The (i, j)-th entry a_{ij} is the number of edges between v_i and v_j; in particular, if G is simple, then a_{ij} is either 1 or 0 depending on whether v_i and v_j are adjacent (1 if $v_i \sim v_j$, 0 otherwise).

Let $\mathbf{e}_k = (\delta_{ik} : i \in \mathbb{N})$ be the the complete orthonormal system in $L_2(G)$ as discussed above. Then the adjacency matrix of a locally-finite graph can be interpreted as a linear operator $\mathbf{A}$ over $L_2(G)$, which is defined on the basis vectors $\mathbf{e}_k$ as follows:

$$\mathbf{A}\mathbf{e}_k = (a_{ik})_{i \in \mathbb{N}},$$

or equivalently

$$\langle \mathbf{A}\mathbf{e}_k, \mathbf{e}_i \rangle = a_{ik}.$$

Since G is locally-finite, $\mathbf{A}$ is well defined, i.e., $\mathbf{A}\mathbf{e}_k$ is an element of $L_2(G)$, and $\mathbf{A}$ can be extended by linearity to a dense subspace of $L_2(G)$, which is

spanned by the basis vectors $\{\mathbf{e}_k : k \in \mathbb{N}\}$. The next theorem gives us a necessary and sufficient condition for the adjacency operator to be continuous or compact.

Theorem 3.5 *[18] The adjacency operator $\boldsymbol{A}$ is bounded if and only if there exists a constant $M < \infty$ such that $d(v) \leq M$ for every vertex $v \in V(G)$. In this case*

$$||A|| \leq M \text{ and } \quad \sigma(\boldsymbol{A}) \subseteq [-M, M].$$

Moreover, $\boldsymbol{A}$ is compact if and only if G has finitely many edges.

If $d(G) = \sup\{d(u)|u \in V\} < \infty$, then $\mathbf{A}$ acts on $L_2(G)$ as a self-adjoint operator. In fact, it is known [18] that uniform boundedness of the vertex degrees in G is sufficient for the adjacency operator to be self-adjoint. For the general case, since $\mathbf{A}$ is real, there is a self-adjoint extension of $\mathbf{A}$ which is not unique in general.

Definition 3.1 *Let G be a locally finite infinite graph. Let $\boldsymbol{D}(G)$ be the diagonal matrix diag$(d(v), v \in V)$ and $\boldsymbol{A} = \boldsymbol{A}(G)$ is the adjacency matrix of G. The Laplacian of G is a linear operator defined by*

$$\boldsymbol{L} = \boldsymbol{D}(G) - \boldsymbol{A}(G). \tag{3.19}$$

Notice that if G is regular, then most of the results for $\mathbf{A}$(G) carry over to $\mathbf{L}$.

Let $\mathcal{D}$ be the set of complex functions on V(G) with finite support. i.e.,

$$\mathcal{D} = \{f : V(G) \to \mathbb{C} \ : \ |\text{supp}(f)| < \infty\}. \tag{3.20}$$

$\mathcal{D} \subset L_2(G)$, is dense in $L_2(G)$. Furthermore, it is known [78, 79] that the Laplacian operator with domain $\mathcal{D}$ is essentially self-adjoint.

For any $f \in \mathcal{D}$ we have

$$\mathbf{L}f(x) = d(x)f(x) - \sum_{y:y\sim x} f(y) = \sum_{y:y\sim x} \left(f(x) - f(y)\right). \tag{3.21}$$

Its quadratic form

$$\langle f, \mathbf{L}f \rangle = \sum_{x \sim y} (f(x) - f(y))^2 \geq 0.$$

Notice that the Laplacian operator, $\mathbf{L}$, is densely defined in the Hilbert space $L_2(G)$ of all square summable functions on the vertices of G; and $\mathbf{L}$ is not defined everywhere in $L_2(G)$ but rather it has a dense domain $\mathcal{D}$ in $L_2(G)$. Moreover, unless the degrees of the vertices are uniformly bounded, the Laplacian is generally an unbounded operator with dense domain.

Theorem 3.6 *[78] The Laplacian operator $\boldsymbol{L}$ with domain $\mathcal{D}$ is essentially self-adjoint.*

Theorem 3.6 means that $\mathbf{L}$ has a unique self-adjoint extension. Furthermore, if $d(G) < \infty$ (i.e., if the vertices of G are uniformly bounded), then by theorem (3.5), $\mathbf{L}$ is a bounded operator as both $\mathbf{D}(G)$ and $\mathbf{A}$(G) are bounded operators. The result can be stated and proved as follows.

Theorem 3.7 *Let $G = (V, E)$ denote a locally finite connected graph. Then the Laplacian $\boldsymbol{L}$ is a bounded operator on $L_2(G)$ if and only if the degree of G is bounded, i.e.,*

$$d(G) = \sup_{v \in V} d(v) < \infty.$$

Proof. If the degree $d(G)$ is bounded from above, a simple calculation using the triangle inequality and the Cauchy-Schwarz inequality leads to
$||\mathbf{L}|| \leq 2d(G)$.
On the other hand, if $d(G)$ is unbounded we choose a sequence $(v_j)_{j \in \mathbb{N}}$ in V with $\sup_{j \in \mathbb{N}} d(v_j) = \infty$ and define $f_j : V \to \mathbb{C}$ by $f_j(v_j) = 1$ and $f_j(v) = 0$ if $v \neq v_j$. Then we clearly have $f_j \in \mathcal{D}$ and

$$(\mathbf{L}f_j)(v) = \begin{cases} d(v_j) & \text{if } v = v_j \ , \\ -1 & \text{if } v \sim v_j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $||\mathbf{L}f_j||^2 = d(v_j)^2 + d(v_j)$ is unbounded but $||f_j|| = 1$. ■

The essential self-adjointness of the Laplacian was also independently proved in [79] using the maximum principle for subharmonic functions, stated as follows:

The maximum principle for subharmonic functions: Let $G = (V, E)$ be a connected graph and let $f : V(G) \to \mathbb{R}$ satisfy $\mathbf{L}f \leq 0$ and assume that there is a vertex $v \in V$ with $f(v) = \max\{f(u) : u \in V\}$. Then f is constant. The proof of the maximum principle is not difficult. In fact, since $\mathbf{L}f(v) \leq 0$ and from eq.(3.21), it follows immediately that

$$0 \geq \mathbf{L}f(v) = \sum_{u \sim v} (f(v) - f(u)) \geq 0$$

since f attains its maximum at v. It follows that $f(u) = f(v)$ for all $u \sim v$. Since the graph is assumed to be connected, the result follows by induction.

3.6 Directed graphs

Directed Graph: A *directed graph*, or *digraph*, G, consists of a set of vertices V(G), a set of edges E(G), and a function which assigns each edge e an ordered pair of vertices (u, v). We call u the tail of e, v the head of e, and u, v the ends of e. If there is an edge with tail u and head v, then we let (u, v) denote such an edge, and we say that this edge is directed from u to v. An edge $e = (u, v)$ in a digraph G is a *loop* if $u = v$. Two edges e & f are *parallel* if they have the same tails and the same heads. If G has no loops or parallel edges, then we say G is *simple*. As in the undirected graphs, it is helpful to represent them with drawings so that each vertex corresponds to a distinct point, and each edge from u to v is represented by a curve directed from the point corresponding to u to the point corresponding to v (usually we indicate this direction with an arrow head). If G is a directed graph, then there is an ordinary (undirected) graph $\widetilde{G}$ with the same vertex and edges as G which is

obtained from G by associating edge (u, v) with ends u, v (in other words, we just ignore the directions of the edges). We call $\widetilde{G}$ the underlying (undirected) graph, and we call G an orientation of $\widetilde{G}$.

Unlike undirected graphs, a vertex of a digraph has two degrees. The *outdegree* of a vertex v, denoted $d^+(u)$ is the number of edges with tail v and the *indegree* of v, denoted $d^-(v)$ is the number of edges with head v. For the sake of convenience, it is customary to denote the outdegree by $d(u)$ instead of $d^+(u)$. It is easy to see that the sum of the indegrees of vertices of a digraph equals the sum of outdegrees. That is

$$\sum_{v \in V(G)} d(v) = |EG| = \sum_{v \in V(G)} d^-(v).$$

A digraph is *regular* provided the indegree and outdegree of each vertex is the same, say some constant k. In this case we say G is k-regular.

The notions of subgraph, spanning subgraph, and induced subgraph are precisely analogous to those for undirected graphs. Connectivity in undirected graphs is straightforward. In a digraph however, connectivity is more subtle. A directed walk in a digraph G is a sequence $v_0, e_1, v_1, \ldots, e_n, v_n$ so that $v_i \in V(G)$ for every $0 \leq i \leq n$, and so that e_i is an edge from v_{i-1} to v_i for every $1 \leq i \leq n$. We say that this is a walk from v_0 to v_n. If $v_0 = v_n$ we say the walk is *closed* and if $v_0, v_1, \ldots, v_n$ are distinct we call it a *directed path*. So a path is an open walk with no repeated vertices or edges. Obviously, if there is a directed walk from u to v, then there is a directed path from u to v. In fact, every directed walk from u to v of minimum length is a directed path. A closed path (*cycle*) is a closed walk with no repeated vertices or edges except that $v_0 = v_n$. A directed *acyclic graph* (DAG) is a directed graph that has no cycles of any length, i.e., for any vertex u there is no directed path that ends with itself.

We say two vertices u and v of a digraph G = (V, E) are connected if there is a (directed) path from u to v and one from v to u. This relation between vertices is reflexive, symmetric, and transitive, so it is an equivalent relation on the vertices. As such, it partitions V into disjoint sets called *strongly connected*

components of the digraph. A digraph G is *strongly connected* provided that for each pair of distinct vertices u and v, there is a directed path from u to v and one from v to u. A digraph G is *weakly connected* if its underlying graph is connected. An undirected graph can be viewed as a digraph with directed edges (u, v), and (v, u) for each undirected edge (u, v).

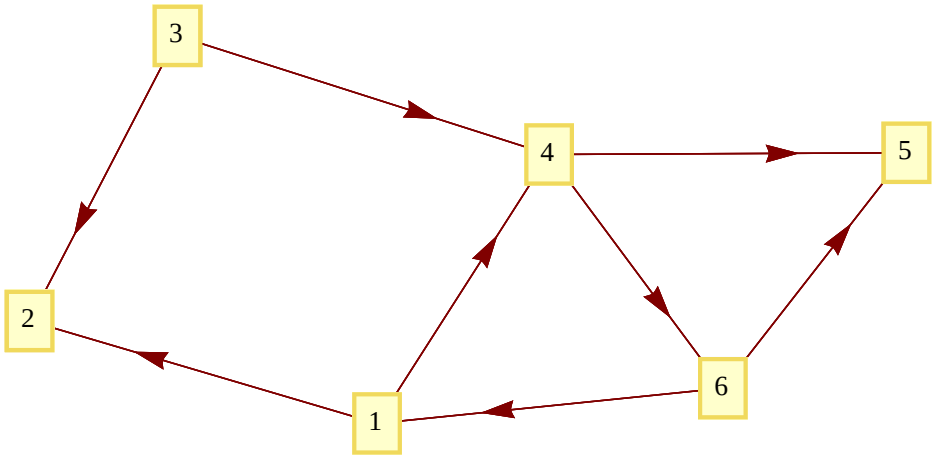


Figure 3.1: A weakly connected directed graph

Define $u \equiv v$ provided there is a walk from u to v and a walk from v to u. This is an equivalent relation and thus V is partitioned into equivalence classes $V_1, V_2, \ldots V_k$. The k subgraphs induced on the sets of vertices $V_1, V_2, \ldots, V_k$ are the strong components of G. A digraph G is strongly connected if and only if it has exactly one strong component. The following theorem summarizes some important properties concerning these notions.

Theorem 3.8 *[63] Let G = (V, E) be a digraph.*

1. *Then G is strongly connected if and only if there does not exist a partition of V into two nonempty sets X and Y such that all the edges between X and Y have their initial vertex in X and their terminal vertex in Y.*

2. *The strong components of G can be ordered as $G_1, G_2, \ldots, G_k$ so that if (u, v) is an edge with u in G_i and v in G_j with $i \neq j$, then $i < j$ (in*

the ordering $G_1, G_2, \ldots, G_k$ all edges between the strong components go from left to right).

Graphs arising from many applications such as transportation problems in which traffic flow is restricted to one direction, one-way communication problems, asymmetric social interactions, athletic tournaments, learning theory, scheduling, flights and even the world-wide web are directed, where direction of links contains crucial information. In many problems, such as traffic problems with one way streets, we must restrict the direction of movement along the edges to only one direction. Many applications in communication networks and in approximation algorithms also involve directed graphs. Despite the fact that there is a large amount of problems which can be naturally represented as directed graphs, the spectral approach for directed graphs has not been as well developed.

The graph Laplacian operator plays a significant role in developing the theory of sampling and approximation on graphs. For undirected graphs, the graph Laplacian operator is a well studied subject. There is a large literature on the spectrum of the Laplacian for such graphs. In addition to its mathematical importance, the spectrum of the Laplace operator has various applications in many fields, as we have mentioned in the previous chapters. Contrary to their huge possible applications and natural existence of directed graphs, very few research has been done in this particular area. The directed version of a Laplace operator was developed recently by Fan Chung [11]. Some properties of its spectrum were investigated. For example, it has been shown that the directed Laplacian operator is self adjoint on the space of functions $L_2(G)$, its eigenvalues are contained in [0,2] (see [11] for details). Despite in its early stage of development, several papers have started to emerge applying the directed Laplacian operator and its spectrum [12, 21, 58, 61, 64, 69, 70, 71].

The main challenge of extending the sampling theory to directed graphs is the complex nature of directed graphs and the study of spectral theory on directed graphs is under development and little information is available in

current literature. Following Chung's approach, a few other researchers have introduced different types of directed and mixed Laplacian operators (see for example [61, 68]). The Laplace operator of a directed graph developed by Chung has received a considerable amount of attention and it turns out that it is extremely useful and has gained many applications in different fields and we will adopt this definition of directed Laplacian in this work. We will show that the directed Laplacian is a generalization of its underlying graph (the graph obtained by ignoring all the directions of the edges). Furthermore, in the case of regular directed graphs, the directed Laplacian and the Laplacian of the underlying graph happens to be the same (we will show this as well). The directed Laplacian operator introduced in [61] is not self-adjoint, which makes it less attractive for many researchers in the literature because working with non-symmetric operators is a bit awkward and inconvenient though they are also useful for many applications. In [68], Bapat, Grossman and Kulkarani introduced a Laplacian operator for mixed graphs as follows:

Given a mixed graph G on vertices $1, \ldots, n$, write $ij \in E(G)$ to mean the existence of an undirected edge between the vertices i and j. Write (i, j) to mean the existence of the directed edge from the vertex i to the vertex j. The adjacency matrix $A(G) = (a_{ij})$ of G is the matrix with

$$a_{ij} = \begin{cases} 1 & \text{if } ij \in E(G) \ , \\ -1 & \text{if } (i,j) \in E(G) \ , \\ 0 & \text{otherwise.} \end{cases} \tag{3.22}$$

Given a mixed graph G, the degree $d(i)$ of a vertex i is the number of edges (both directed and undirected) incident with i. Let D(G) be the diagonal matrix with $d(i)$ as the i-th diagonal entry. The *Laplacian matrix* L(G) is defined as the matrix

$$L(G) = D(G) + A(G).$$

This matrix is shown to be positive semi-definite.

The Laplacian spectrum of mixed graphs have been studied since then and

further investigated in [68, 69, 70, 71].

For example, Zhang and Li [70], and Zhang and Luo [71] gave some upper bounds for the spectral radius and the second smallest eigenvalue of the Laplacian matrix of a mixed graph.

Mixed graphs are very important for the study of graph theory as they provide a setup where one can have both directed and undirected edges in the graph. Many real world data structures can be modeled by mixed graphs. For example, streets in a particular city and airline routes in which some streets and airway routes are only one way and some are both ways, can be represented by mixed graphs. So it makes sense to consider those graphs and study their spectral properties.

3.7 Laplacian on directed graphs

Given a digraph G, there is a natural random walk on G with the transition probability function $p : V \times V \to \mathbb{R}^+$ defined by

$$p(u,v) = \begin{cases} \frac{1}{d(u)} & \text{if}(u,v) \in E(G); \\ 0 & \text{otherwise.} \end{cases} \tag{3.23}$$

Note that $p(u,v)$ denotes the probability of moving from vertex u to vertex v. Moreover, as in the undirected case, for each vertex $u \in V$,

$$\sum_{v \in V} p(u,v) = 1. \tag{3.24}$$

However, in the case of digraphs, it is not always true that $\sum_{u \in V} p(u,v) = 1$.

The transition probability matrix associated to G is the matrix P whose $(i,j)^{th}$ is $p(i,j)$. P is not symmetric in general. The Perron-Frobenius Theorem implies that the transition probability matrix P of a strongly connected graph has a unique left eigenvector ϕ with $\phi(v) > 0$ for all v, and $\phi P = \phi$. That is, the random walk on a strongly connected directed graph has a unique

stationary ϕ satisfying the equation

$$\phi(v) = \sum_{u \to v} \phi(u) p(u, v), \tag{3.25}$$

for all $v \in V$, and $\phi(v) > 0$ for all $v \in V$. If G is strongly connected and aperiodic, the random walk converges to the stationary distribution (also known as the Perron vector) ϕ, [11]. The transition probability matrix and Perron vector will be discussed in more detail later.

A digraph is *weighted* when there is a function $w : E \to \mathbb{R}^+$ which associates a positive value $w(u, v)$ with each $(u, v) \in E$. The function w is called a *weight function*. Generally, it is always possible to equip a graph with a canonical weight function defined by $w(u, v) = 1$ at each edge $(u, v) \in E$. In this case the indegree and the outdegree of a vertex v are given by

$$w^-(v) = \sum_{u:\, u \to v} w(u, v) = \sum_{u \in V} w(u, v), \tag{3.26}$$

and

$$w^+(v) = w(v) = \sum_{u:\, v \to u} w(v, u) = \sum_{u \in V} w(v, u). \tag{3.27}$$

For a weighted digraph with edge weights $w(u, v) \geq 0$, a transition probability function p can be defined as

$$p(u, v) = \frac{w(u, v)}{\sum_{z \in V} w(u, z)} = \frac{w(u, v)}{w(u)}. \tag{3.28}$$

Order the vertices of G in some way: $v_1, v_2, \ldots, v_n$. The *adjacency matrix* of a digraph G with n vertices is the $n \times n$ matrix $A = (a_{ij})$, where $a_{ij} = 1$ if there is a directed edge from i to j and 0 otherwise. A different ordering results in the (similar) matrix BAB^T for some permutation matrix B. In particular, the digraph is strongly connected if and only if the matrix A is *irreducible*. If G is a weighted digraph, then a_{ij} is replaced by $w(i, j)$, the weight of edge (i, j). Notice that unlike the adjacency matrix of an undirected graph the adjacency matrix of a digraph is no longer symmetric. We also have the relation $P = D^{-1}A$, as in the undirected case, where D = diag$(d(v), v \in V)$

= diag(A.**1**) is the degree matrix (the diagonal matrix with diagonal entries $D(v, v) = d(v)$, the out degree of v).

Let Φ be an $n \times n$ diagonal matrix with diagonal entries $\Phi(v, v) = \phi(v)$, i.e.,

$$\Phi = \begin{pmatrix} \phi(v_1) & 0 & \cdots & 0 \\ 0 & \phi(v_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi(v_n) \end{pmatrix}. \tag{3.29}$$

Φ is invertible and non-negative since $\phi(v) > 0$, $\forall v \in V$. The directed Laplacian matrix of a digraph G is defined in terms of P and Φ as follows.

Definition 3.2 *Let $G = (V, E)$ be a digraph with $|V(G)| = n$. Let P be the transition probability matrix associated to G and Φ be the $n \times n$ matrix defined as in eq.(3.29). The normalized Laplacian of G is defined as*

$$\mathcal{L} = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^* \Phi^{1/2}}{2}, \tag{3.30}$$

where I is the identity matrix and P^* denotes the transpose of P.

For undirected graph $G = (V, E)$, it is well known [9] that the Perron vector of the natural random walk has a closed-form expression

$$\phi(v) = \frac{d(v)}{\sum_u d(u)} = \frac{d(v)}{\text{vol(G)}}, \tag{3.31}$$

where $\text{vol(G)} = \sum_{u \in V} d(u)$.

Using equation eq.(3.31), we will show that eq.(3.30) reduces to the well known Laplacian matrix for undirected graphs. It is also easy to see that the directed Laplacian satisfies $\mathcal{L} = \mathcal{L}^*$, that is $\mathcal{L}$ is a self-adjoint operator.

The unnormalized directed Laplacian L of a directed graph is also defined as

$$L = \Phi - \frac{\Phi P + P^* \Phi}{2}, \tag{3.32}$$

where P and Φ are as above. Now, the normalized Laplacian $\mathcal{L}$ can be written in terms of the unnormalized Laplacian L as follows

$$\mathcal{L} = \Phi^{-1/2} L \Phi^{-1/2}. \tag{3.33}$$

Let $L_2(G)$ denote the space of all real valued functions $f : V(G) \to \mathbb{R}$ which assigns a value $f(v)$ for each vertex v. A function in $L_2(G)$ can be thought of as a column vector in $\mathbb{R}^n$, where $n = |V(G)|$, the number of vertices in G. The space $L_2(G)$ then can be endowed with the standard inner product in $\mathbb{R}^n$ as

$$\langle f, g \rangle_{L_2(G)} = \sum_{v \in V} f(v) g(v) \tag{3.34}$$

for all $f, g \in L_2(G)$ and the norm

$$||f||_{L_2(G)} = ||f|| = \left(\sum_{v \in V} |f(v)|^2 \right)^{1/2}. \tag{3.35}$$

For any $f \in L_2(G)$ we have the following explicit form

$$(\mathcal{L}f)(v) = f(v) - \frac{1}{2} \left(\sum_{u,\, u \to v} \frac{\phi(u) p(u,v) f(u)}{\sqrt{\phi(u)\phi(v)}} + \sum_{w,\, v \to w} \frac{\phi(v) p(v,w) f(w)}{\sqrt{\phi(w)\phi(v)}} \right). \tag{3.36}$$

Now let's show that eq.(3.36) reduces to eq.(3.4), the well known formula for undirected graph Laplacian

$$(\mathcal{L}f)(v) = \frac{1}{\sqrt{d(v)}} \sum_{v \sim u} \left(\frac{f(v)}{\sqrt{d(v)}} - \frac{f(u)}{\sqrt{d(u)}} \right), \quad f \in L_2(G), \tag{3.37}$$

where $v \sim u$ means that v is adjacent to u. To see this, we recall from eq.(3.31)

that the Perron vector $\phi = \left(\frac{d(v)}{vol(G)}\right)_{v\in V}$ for undirected graphs. So we have

$$\begin{aligned}
(\mathcal{L}f)(v) &= f(v) - \frac{1}{2}\left(\sum_{u\to v}\frac{\phi(u)\,p(u,v)\,f(u)}{\sqrt{\phi(u)\,\phi(v)}} + \sum_{v\to u}\frac{\phi(v)\,p(v,u)f(u)}{\sqrt{\phi(u)\,\phi(v)}}\right)\\
&= f(v) - \frac{1}{2}\sum_{u\sim v}\frac{(d(u)/vol(G))\,(1/d(u))\,f(u)}{\sqrt{(d(u)/vol(G))\,(d(v)/vol(G))}}\\
&- \frac{1}{2}\sum_{v\sim u}\frac{(d(v)/vol(G))\,(1/d(v))\,f(u)}{\sqrt{d(u)/vol(G)\,d(v)/vol(G)}}\\
&= f(v) - \frac{1}{2}\left(\sum_{u\sim v}\frac{f(u)}{\sqrt{d(u)\,d(v)}} + \sum_{v\sim u}\frac{f(u)}{\sqrt{d(u)\,d(v)}}\right)\\
&= f(v) - \sum_{u\sim v}\frac{f(u)}{\sqrt{d(v)\,d(u)}}\\
&= \sum_{u\sim v}\frac{f(v)}{d(v)} - \frac{1}{\sqrt{d(v)}}\sum_{u\sim v}\frac{f(u)}{\sqrt{d(u)}}\\
&= \frac{1}{\sqrt{d(v)}}\sum_{u\sim v}\left(\frac{f(v)}{\sqrt{d(v)}} - \frac{f(u)}{\sqrt{d(u)}}\right),
\end{aligned}$$

as desired. ■

There is also a natural way to connect the Laplacian of a directed graph with the Laplacian of the corresponding undirected graph. Let us first prove the following lemma from [21].

Lemma 3.1 *If G is a directed graph and ϕ is the Perron vector of the transition probability matrix P of G, then*

$$\sum_{\substack{u\\ u\to v}}\phi(u)p(u,v) = \sum_{\substack{w\\ v\to w}}\phi(v)p(v,w). \tag{3.38}$$

Proof: The proof is straightforward by recalling eq.(3.24) that $\sum_w p(v,w) = 1$ for each vertex v and using the fact that ϕ is the (left) eigenvector of P with eigenvalue 1. So we have $\phi\,P = \phi$ and hence

$$\phi(v) = \sum_{\substack{u\\ u\to v}}\phi(u)p(u,v).$$

Thus,

$$\sum_{\substack{u \\ u\to v}} \phi(u)p(u,v) = \phi(v) = \phi(v)\sum_{\substack{w \\ v\to w}} p(v,w) = \sum_{\substack{w \\ v\to w}} \phi(v)p(v,w)$$

■

Theorem 3.9 *Let G be an aperiodic strongly connected weighted graph and let $\widetilde{G}$ be the underlying graph (the graph obtained by disregarding the direction on the edges of G) with weights defined by*

$$w(u,v) = \phi(u)p(u,v) + \phi(v)p(v,u). \tag{3.39}$$

Then $\mathcal{L}(G) = \mathcal{L}(\widetilde{G})$.

Proof: Let $w_G(v)$ and $w_{\widetilde{G}}(v)$ denote the weighted degrees of the vertex v in G and $\widetilde{G}$ respectively. Then it follows from Lemma 3.1, equations (3.24) and (3.25) that

$$\begin{aligned}
w_{\widetilde{G}}(v) &= \sum_u w(u,v) = \sum_u (\phi(u)p(u,v) + \phi(v)p(v,u)) \\
&= \sum_u \phi(u)p(u,v) + \sum_u \phi(v)p(v,u) \\
&= \phi(v)\underbrace{\sum_u p(v,u)}_{=1 \text{ by } (3.24)} + \sum_u \phi(u)p(u,v) \\
&= \phi(v) + \underbrace{\sum_u \phi(u)p(u,v)}_{=\phi(v) \text{ by } (3.25)} \\
&= 2\phi(v).
\end{aligned}$$

On the other hand, for each $v \in V(G)$ we have (follows from eq.(3.30)),

$$\mathcal{L}(G)(v,v) = 1 - p(v,v) = 1 = \mathcal{L}(\widetilde{G})(v,v), \text{ since } p(v,v) = 0.$$

Thus, $\mathcal{L}(G)$ and $\mathcal{L}(\widetilde{G})$ have the same diagonal entries.

Now for $u, v \in V(G), u \neq v$,

$$\begin{aligned}
\mathcal{L}(G)(u,v) &= -\frac{1}{2}\left(\sqrt{\frac{\phi(u)}{\phi(v)}}p(u,v) + \sqrt{\frac{\phi(v)}{\phi(u)}}p(v,u)\right) \\
&= -\frac{(\phi(u)p(u,v) + \phi(v)p(v,u)}{\sqrt{(2\phi(u)(2\phi(v))}} \\
&= -\frac{w(u,v)}{\sqrt{w_G(u)\, w_G(v)}} \\
&= \mathcal{L}(\widetilde{G})(u,v),
\end{aligned}$$

hence they have the same off-diagonal elements as well. This completes the proof. ■

One can also verify that the quadratic form of the directed Laplacian is given by

$$\langle f, \mathcal{L}f\rangle = \frac{1}{2}\sum_{(u,v)\in E}\phi(u)P(u,v)\left(\frac{f(u)}{\sqrt{\phi(u)}} - \frac{f(v)}{\sqrt{\phi(v)}}\right)^2 \geq 0, \tag{3.40}$$

which shows that the Laplacian operator $\mathcal{L}$ is semi-positive definite.

As in the case of undirected graphs, the eigenvalues of the direct Laplacian are found to be useful. Below, we will show that for any strongly connected digraph G, the spectrum of $\mathcal{L}$ is contained in [0,2].

Let $\sigma(\mathcal{L})$ denote the spectrum of $\mathcal{L}$, and let

$$Q = \frac{\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^*\Phi^{1/2}}{2}.$$

The following theorem shows that the eigenvalues of Q are bounded between -1 and 1.

Theorem 3.10 *If Q is defined as above, then $\sigma(Q) \subseteq [-1,1]$ and $\sqrt{\phi}$ is an eigenvector of Q with corresponding eigenvalue 1.*

Proof: Let R be the matrix defined by $R = \frac{P+\Phi^{-1}P^*\Phi}{2}$. R is an operator in

$L_2(G)$. Also note that

$$\begin{aligned}\Phi^{1/2}R\Phi^{-1/2} &= \Phi^{1/2}\left(\frac{P+\Phi^{-1}P^*\Phi}{2}\right)\Phi^{-1/2}\\ &= \frac{\Phi^{1/2}P\Phi^{-1/2}+\Phi^{-1/2}P^*\Phi^{1/2}}{2}\\ &= Q.\end{aligned}$$

Thus, R and Q are similar and hence they have the same set of eigenvalues. Suppose f is an eigenvector of R with eigenvalue λ. Choose v such that $|f(v)|$ $= \max_{u\in V}|f(u)|$. Then it suffices to show that $|\lambda| \leq 1$.

$$\begin{aligned}|\lambda||f(v)| &= \left|\sum_{u\in V} R(v,u)\, f(u)\right|\\ &\leq \sum_{u\in V} R(v,u)|f(v)|\\ &= \frac{|f(v)|}{2}\left(\sum_{v\to u} p(v,u) + \sum_{u\to v}\frac{\phi(u)p(u,v)}{\phi(v)}\right)\\ &= \frac{|f(v)|}{2}\left(\sum_{v\to u} p(v,u) + \sum_{v\to w}\frac{\phi(v)p(v,w)}{\phi(v)}\right), \qquad \text{(by Lemma 3.1)}\\ &= \frac{|f(v)|}{2}\left(2\sum_{v\to u} p(v,u)\right)\\ &= |f(v)|,\end{aligned}$$

which implies $|\lambda| \leq 1$.

Now it remains to show that $Q\sqrt{\phi} = \sqrt{\phi}$. To see this, first note that $Q = I - \mathcal{L}$.

Then the result follows from (3.36) since for any $v \in V$,

$$\begin{aligned}
(Q\sqrt{\phi})(v) &= [(I-\mathcal{L})\sqrt{\phi}](v) = \sqrt{\phi(v)} - (\mathcal{L}\sqrt{\phi})(v) \\
&= \frac{1}{2}\left(\sum_{u\to v} \frac{\phi(u)p(u,v)\sqrt{\phi(u)}}{\sqrt{\phi(u)\,\phi(v)}} + \sum_{v\to u} \frac{\phi(v)p(v,u)\sqrt{\phi(u)}}{\sqrt{\phi(u)\,\phi(v)}}\right) \\
&= \frac{1}{2}\left(\sum_{u\to v} \frac{\phi(u)p(u,v)}{\sqrt{\phi(v)}} + \sum_{v\to u} \frac{\phi(v)p(v,u)}{\sqrt{\phi(v)}}\right) \\
&= \frac{1}{2}\left(\frac{1}{\sqrt{\phi(v)}} \underbrace{\sum_{u\to v} \phi(u)p(u,v)}_{=\phi(v)} + \sqrt{\phi(v)} \underbrace{\sum_{v\to u} p(v,u)}_{=1}\right) \\
&= \sqrt{\phi(v)}.
\end{aligned}$$

The next corollary follows from the previous theorem and noting that $\mathcal{L} = I - Q$.

Corollary 3.4 *Let G be a strongly connected digraph and $\mathcal{L}$ is its directed Laplacian matrix. Then $\sigma(\mathcal{L}) \subseteq [0, 2]$. Furthermore, 0 is an eigenvalue of $\mathcal{L}$ with eigenvector $\sqrt{\phi}$.*

Remark 3.2 *The definition of the directed Laplacian depends on the transition probability matrix and is restricted to strongly connected and aperiodic graphs only so that the natural random walk on the graph converges to a unique positive stationary distribution. Obviously this is too restrictive and most graphs in real world application are neither strongly connected nor aperiodic. To overcome this restriction we may invoke the so called "**teleporting random walk** " on the graph and extend the definition of the Laplacian to general directed graphs. But first we will review the Perron-Frobenius Theorem and its significance on directed graphs.*

3.8 A brief review of the Perron-Frobenius Theorem

We have seen that the transition probability matrix P of a directed graph is not symmetric in general. However, it has the useful property that all of its entries are non-negative. The Perron-Frobenius Theorem gives useful information about the eigenvalues of such matrices in many cases.

Definition 3.3 *A real matrix $M = (m_{ij})$ is called nonnegative if each of its entries is non-negative. A nonnegative, square matrix M is* irreducible *if for each pair (i, j), there is a nonnegative integer k such that the $(i, j)^{th}$ entry of M^k is strictly positive.*

Since the $(i, j)^{th}$ entry of the k^{th} power of the adjacency matrix of a connected undirected graph G is equal to the number of edge sequences of length k connecting vertex i to vertex j, it is always irreducible. The following theorem provides an important algebraic characterization of irreducible, nonnegative matrices.

Theorem 3.11 *[62] Let M be a nonnegative matrix of order n. Then M is irreducible if and only if $(I + M)^{n-1}$ is a positive matrix.*

For each i, the row sum r_i of a matrix $M = (m_{ij})$ is given by

$$r_i = \sum_j m_{ij}.$$

Theorem 3.12 *([Perron-Frobenius]) Let M be a nonnegartive, square matrix, and suppose M is irreducible. Let r_{min} and r_{max} be the minimum and maximum row sums of M, respectively. There is a unique eigenvector v of M all of whose entries are positive. The eigenvalue λ corresponding to v is the largest eigenvalue of M and satisfies*

$$r_{min} \leq \lambda \leq r_{max}$$

Theorem 3.13 *[62] Let A be a matrix with nonnegative entries. Assume further that A is irreducible, equivalently meaning that it is impossible to permute the rows and columns of A to write it in the form*

$$\tilde{A} = \begin{pmatrix} X & 0 \\ Y & Z \end{pmatrix}$$

with the upper right block having dimension $k \times (n-k)$ for some k. Then there is a real $\rho_0 > 0$ such that the following hold:

1. *ρ_0 is an eigenvalue of A, and all other eigenvalues ρ_i satisfy $|\rho_i| \leq \rho_0$.*
2. *The eigenvector (which is unique) corresponding to the eigenvalue ρ_0 has all entries positive.*
3. *If there are $k-1$ other eigenvalues with $|\rho_i| = \rho_0$, then they are of the form $\rho_0 \theta^j$, where $\theta = e^{\frac{2\pi i}{k}}$.*

Now let us translate Perron-Frobenius Theorem into graphs. Notice that in order to apply Perron-Frobenius theorem to a transition probability matrix on a digraph, we need to make sure that the matrix we apply is irreducible. With this in mind, we have the following theorem:

Theorem 3.14 *A digraph G is strongly connected if any of the following equivalent conditions hold:*

1. *For every u and v in $V(G)$ there exist directed paths in G from u to v and from v to u.*
2. *For any partition of $V(G)$ into two disjoint nonempty sets X and Y there is an edge from X to Y and an edge from Y to X.*
3. *The adjacency matrix A of G is irreducible.*
4. *The transition probability matrix P of G is irreducible.*

Combining Theorems 3.12, 3.13 and 3.14 we have the following theorem.

Theorem 3.15 *The transition probability matrix P of a strongly connected digraph has a unique left eigenvector ϕ such that $\phi(v) > 0$ for all $v \in V(G)$, and*

$$\phi P = \rho\phi,$$

where ρ is the spectral radius of P.

We will treat ϕ as a row vector. Since $P\mathbf{1} = \mathbf{1}$, we have $\rho = 1$ and the Perron-Frobenius Theorem implies that all eigenvalues of P have absolute value bounded by 1. Remember that P is not symmetric and hence its eigenvalues may not be real. We scale (normalize) and choose ϕ such that

$$\phi\mathbf{1} = \sum_v \phi(v) = 1. \tag{3.41}$$

We call ϕ the *Perron vector* of P.

Definition 3.4 *A strongly connected graph G is called periodic if any of the following equivalent conditions (definitions) hold:*

1. *the transition matrix P has eigenvalue $\rho \neq 1$ such that $|\rho| = 1$.*
2. *There is a $k > 1$ such that all eigenvalues with norm 1 are of the form $e^{\frac{2\pi ij}{k}}$.*
3. *There is an edge preserving map from V(G) to the vertices of the cycle C_k with $k > 1$ (a map such that $u \to v$ in G implies $f(u) \to f(v)$ in C_k).*
4. *The GCD of all cycle lengths in G is $k > 1$.*

We say G is *aperiodic* if it is not periodic. If G is strongly connected and aperiodic, the random walk converges to the Perron vector ϕ.

3.9 Comparing Laplacians on regular graphs

Due to their symmetry, regular graphs are easy to understand and have many interesting applications. For example, in computer graphics, the shape of a 3D object can be approximated by regular graph, with its nodes containing the coordinate information. In the following theorem, we will show that a strongly connected regular digraph and its underlying graph have the same Laplacian.

Theorem 3.16 *Suppose G is a strongly connected k-regular digraph and let $\widetilde{G}$ be its underlying graph. Let $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ be the Laplacians of G and $\widetilde{G}$ respectively. Then $\mathcal{L} = \widetilde{\mathcal{L}}$.*

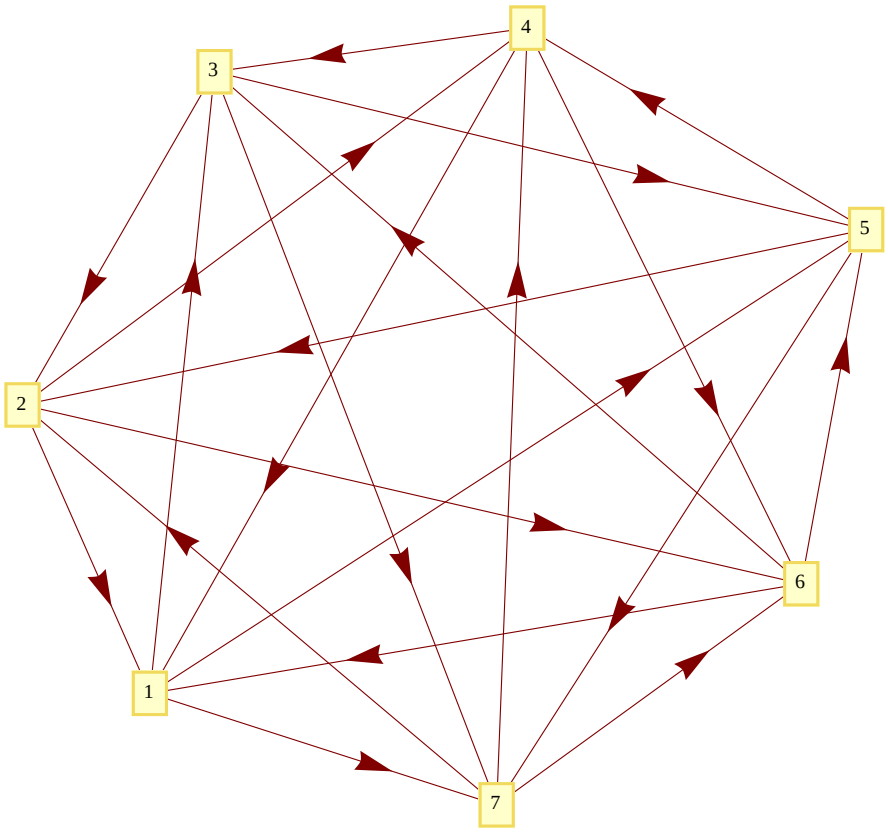
Proof: We know that for a strongly connected regular digraph of degree k, $\phi(v) = \frac{1}{n}$ and $p(u,v) = \frac{1}{k}$, for each $u, v \in V$, where $n = |V|$. Since G is k-regular every vertex $v \in V$ has equal indegree and outdegree k. So $\widetilde{G}$ is $2k$-regular and the Laplacian of $\widetilde{G}$ is given by $\widetilde{\mathcal{L}} = I - \frac{1}{2k}A$, where A is the adjacency matrix of $\widetilde{G}$ and I is the identity matrix. Therefore, for each $v \in V(G)$ and for each $f \in L_2(G)$

$$(\widetilde{\mathcal{L}}f)(v) = f(v) - \frac{1}{2k}\sum_{u:\, u\sim v} f(u).$$

On the other hand, for each $v \in V(G)$, we have

$$\begin{aligned}
(\mathcal{L}f)(v) &= f(v) - \frac{1}{2}\left(\sum_{u:\, u\to v} \frac{\phi(u)\, p(u,v)\, f(u)}{\sqrt{\phi(u)\,\phi(v)}} + \sum_{w:\, v\to w} \frac{\phi(v)\, p(v,w) f(w)}{\sqrt{\phi(w)\,\phi(v)}}\right) \\
&= f(v) - \frac{1}{2}\left(\sum_{u:\, u\to v} \frac{1}{k} f(u) + \sum_{w:\, v\to w} \frac{1}{k} f(w)\right) \\
&= f(v) - \frac{1}{2k}\left(\sum_{u:\, u\to v} f(u) + \sum_{w:\, v\to w} f(w)\right) \\
&= f(v) - \frac{1}{2k}\sum_{u:\, u\sim v} f(u) \\
&= (\widetilde{\mathcal{L}}f)(v).
\end{aligned}$$

This completes the proof. ■



5

Figure 3.2: A strongly connected 3-regular directed graph

Example 1: Consider G, a directed cycle of length 5 and denote the vertex set by $V = \{1, 2, 3, 4, 5\}$ and directed edges $E = \{(1, 2), (2, 3), (3, 4), (4, 5), (5, 1)\}$. It is a regular digraph of order $k = 1$. Its underlying graph G is a regular graph of order 2. The transition probability matrix of G is given by

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

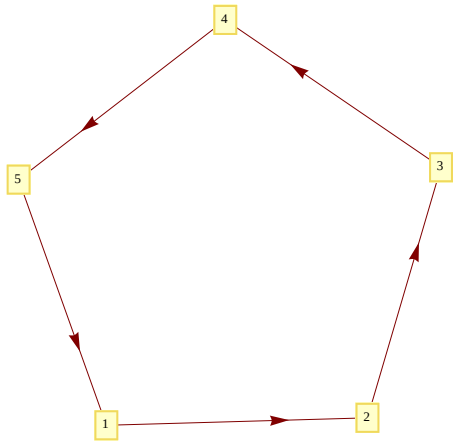


Figure 3.3: directed cycle graph

The eigenvector of P corresponding to the eigenvalue 1 is

$$\phi = (0.4472, 0.4472, 0.4472, 0.4472, 0.4472, 0.4472).$$

Using eq.(3.30), we obtain the directed Laplacian $\mathcal{L}$ of G as follows

$$\mathcal{L} = \begin{pmatrix} 1 & -1/2 & 0 & 0 & -1/2 \\ -1/2 & 1 & -1/2 & 0 & 0 \\ 0 & -1/2 & 1 & -1/2 & 0 \\ 0 & 0 & -1/2 & 1 & -1/2 \\ -1/2 & 0 & 0 & -1/2 & 1 \end{pmatrix}$$

On the other hand, using the definition of the undirected Laplacian we have

$$\widetilde{\mathcal{L}}(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d(v) \neq 0 \ ; \\ -\frac{1}{\sqrt{d(u)d(v)}} & \text{if } u \sim v; \\ 0 & \text{otherwise,} \end{cases}$$

and we obtain

$$\widetilde{\mathcal{L}} = \begin{pmatrix} 1 & -1/2 & 0 & 0 & -1/2 \\ -1/2 & 1 & -1/2 & 0 & 0 \\ 0 & -1/2 & 1 & -1/2 & 0 \\ 0 & 0 & -1/2 & 1 & -1/2 \\ -1/2 & 0 & 0 & -1/2 & 1 \end{pmatrix},$$

which shows $\mathcal{L} = \widetilde{\mathcal{L}}$.

In the case of undirected graphs, it is shown that any finite subset of the vertex set admits a poincaré inequality. In the following theorem we will also show this is generally true for directed graphs. We first construct $\Gamma(S)$ for any subset $S \subset V(G)$ as follows:

Let S be a non-empty subset of the vertex set $V(G)$ and consider the set $\bar{S} = S \cup \partial S$ as an induced subgraph of G. Let $\Gamma(S)$ be a graph constructed in the following way: Take two copies of the induced subgraph $\bar{S}$, and denote them as $\bar{S}_1$ and $\bar{S}_2$, and identify every vertex $v \in \partial S \subset \bar{S}_1$ with the same vertex $v \in \partial S \subset \bar{S}_2$. It is important to note that $\Gamma(S)$ is not generally strongly connected even if G is strongly connected. But for now assume both G and $\Gamma(S)$ are strongly connected and aperiodic. Later we will discuss a remedy to disposeof this strict assumption.

Theorem 3.17 *Suppose $G = (V, E)$ is a strongly connected digraph and S is a nonempty subset of V(G). Then S is a Λ-set with $\Lambda = \frac{\sqrt{2}}{\lambda_1(\Gamma(S))}$.*

That is, for any $\varphi \in L_2(S)$, the inequality

$$\|\varphi\|_{L_2(G)} \leq \frac{\sqrt{2}}{\lambda_1(\Gamma(S))} ||\mathcal{L}\varphi||_{L_2(G)} \tag{3.42}$$

holds true.

Proof: Construct an embedding of the space $L_2(S)$ into the space $L_2(\Gamma(S))$ as follows: If $\varphi \in L_2(S)$, then its image $F_\varphi \in L_2(\Gamma(S))$ is defined by

$$F_\varphi(v) = \begin{cases} \varphi(v) & \text{if } v \in \bar{S}_1, \\ -\varphi(v) & \text{if } v \in \bar{S}_2 \end{cases} \tag{3.43}$$

Then we have

(i)

$$\begin{aligned}||F_\varphi||^2_{L_2(\Gamma(S))} &= \sum_{v\in\Gamma(S)} |F_\varphi(v)|^2 = \sum_{v\in\bar{S}_1} |F_\varphi(v)|^2 + \sum_{v\in S_2} |F_\varphi(v)|^2 \\ &= \sum_{v\in S_1} |\varphi(v)|^2 + \sum_{v\in S_2} |-\varphi(v)|^2 = 2\sum_{v\in S} |\varphi(v)|^2 \\ &= 2\sum_{v\in V(G)} |\varphi(v)|^2 = 2||\varphi||^2_{L_2(G)}.\end{aligned}$$

That is

$$||F_\varphi||_{L_2(\Gamma(S))} = \sqrt{2}||\varphi||_{L_2(G)}. \tag{3.44}$$

(ii) Let $d_{\Gamma(S)}(v)$, $d_{\bar{S}}(v)$, $d_G(v)$ be the outdegrees of v in $\Gamma(S)$, $\bar{S}$ and G respectively. It is easy to see that $d_{\Gamma(S)}(v) \le 2d_G(v)$. As a result

$$||\mathcal{L}_{\Gamma(S)}F_\varphi||_{L_2(\Gamma(S))} \le 2||\mathcal{L}_G\varphi||_{L_2(G)}. \tag{3.45}$$

(iii) From Theorem 3.10, we know that the vector

$$\sqrt{\phi} = \left(\sqrt{\phi(v_1)}, \sqrt{\phi(v_2)}, ..., \sqrt{\phi(v_n)}\right)$$

is an eigenfunction for the Laplacian $\mathcal{L}$ corresponding to the eigenvalue zero.

Thus,

$$\Psi_0 = \left(\sqrt{\phi_{\Gamma(S)}(v)}\right)_{v\in\Gamma(S)} = \left(\sqrt{\phi_{\Gamma(S)}(v_1)}, \sqrt{\phi_{\Gamma(S)}(v_2)}, \ldots, \sqrt{\phi_{\Gamma(S)}(v_N)}\right)$$

is the eigenfunction of $\mathcal{L}_{\Gamma(S)}$ which corresponds to the eigenvalue 0, where $\phi_{\Gamma(S)}$ is the stationary distribution of the natural random walk on $\Gamma(S)$. Since every function F_φ is "odd" (in the sense that $\varphi(v) = -\varphi(v)$ when v is considered as an element in $\bar{S}_2$) it is orthogonal to the subspace spanned by Ψ_0. Now let $\{\Psi_k\}, k = 0, 1, \ldots, N$ is a complete orthonormal system of eigenfunctions of

$\mathcal{L}_{\Gamma(S)}$. (Note: Any real self-adjoint operator has an orthonormal basis of real eigenvectors with its associated eigenvalues). Hence we have

$$F_\varphi = \sum_{j=1}^{N} \langle F_{\varphi,\Psi_j}\rangle \Psi_j,$$

and

$$\mathcal{L}_{\Gamma(S)} F_\varphi = \sum_{j=1}^{N} \lambda_j(\Gamma(S)) \langle F_{\varphi,\Psi_j}\rangle \Psi_j.$$

Thus,

$$||\mathcal{L}_{\Gamma(S)} F_\varphi||^2_{L_2(\Gamma(S))} = \sum_{j=1}^{N} {\lambda_j}^2(\Gamma(S)) |\langle F_{\varphi,\Psi_j}|^2 \geq {\lambda_1}^2(\Gamma(S)) ||F_\varphi||^2_{L_2(\Gamma(S))}.$$

The last inequality follows from the fact that $\lambda_1 \leq \lambda_j,\ j = 2, 3, \ldots, N$ and *Parseval's identity.*

By taking the square root of both sides we obtain the inequality

$$\lambda_1(\Gamma(S)) ||F_\varphi||_{L_2(\Gamma(S))} \leq ||\mathcal{L}_{\Gamma(S)} F_\varphi||_{L_2}(\Gamma(S)). \tag{3.46}$$

Now combining equations (3.44), (3.45) and (3.46) we obtain the desired result

$$||\varphi||_{L_2(G)} \leq \frac{\sqrt{2}}{\lambda_1(\Gamma(S))} ||\mathcal{L}\varphi||_{L_2(G)}. \qquad \blacksquare$$

Let $h_\Gamma(S)$ be the Cheeger constant of $\Gamma(S)$. It is known [11] that

$$2h_\Gamma(S) \geq \lambda_1(\Gamma(S)) \geq \frac{h^2_{\Gamma(S)}}{2}.$$

Thus, we have from the above inequality

$$||\varphi||_{L_2(G)} \leq \frac{2\sqrt{2}}{h^2_{\Gamma(S)}} ||\mathcal{L}\varphi||_{L_2(G)}.$$

3.10 Laplacian on general directed graphs

In the previous sections, we assumed the digraph to be strongly connected and aperiodic so that the natural random walk over the digraph converges

to a unique and positive stationary distribution ϕ. However, this assumption cannot be guaranteed for a general directed graph. To remedy this problem, we may invoke the so-called *teleporting random walk* as the replacement of the natural random walk. The idea of teleporting random walk was first introduced by Brin and Page in 1998 in their PageRank Model [64] and have been extensively studied since then.

Now consider a general directed graph $G = (V, E)$. Given that we are currently at a vertex u:

(a) with probability α choose an outlink uniformly at random and follow the link to the next vertex; or

(b) with probability 1-α jump to a vertex chosen uniformly at random over the entire vertex set excluding itself.

The constant $0 \leq \alpha \leq 1$ is introduced in the case where the current vertex has outdegree zero. Such a random walk is guaranteed to converge to a unique stationary distribution [64]. By adding in an α probability of jumping to a uniform random vertex, we can avoid many of the problems of random walks on directed graphs. First, it ensures that the walk graph (i.e., the random walk and hence the transition probability matrix) is strongly connected. Second, it ensures that the probability of reaching every vertex is at least α/n, and so is not too small.

The choice of an appropriate α varies between 0 and 1 depending on the problem at hand and the structure of the graph. For instance, if G is strongly connected or nearly strongly connected (but $\Gamma(S)$ is not strongly connected) choosing α close to 1 would be a good choice to make the new transition probability matrix close to the original one. It is also known that as α gets close to 1, the rate of convergence of the random walk becomes slow. More precisely, an α much less than 1 guarantees quick convergence but an α close to 1 is better at preserving the information in P.

The transition probability matrix of teleporting random walk can be writ-

ten explicitly as

$$\tilde{P} = \alpha P + (1-\alpha)ee^T/n, \tag{3.47}$$

where $P = D^{-1}A$ is the the probability matrix of the natural random walk, e is the column vector of all ones, n is the order of the matrix (which also equals $|V|$), A is the adjacency matrix of G, and D is the diagonal matrix whose diagonal entries are the outdegrees of the vertices. In fact, the uniform vector $\frac{1}{n}e^T$ can be replaced with a *general probability vector* $\nu > 0$. A probability vector is a vector with non-negative entries that add up to one. When calculating $P = D^{-1}A$, it may happen that a vertex in G has no outdegree, which in turn results a row of P with all entries zero. So proper care must be given to handle this issue. To fix the problem one remedy is to replace all zero rows with $\frac{1}{n}e^T$ (of course any suitable probability vector ν may be used across the row). To give a brief example on how to construct such $\tilde{P}$, let us consider the following directed graph (see figure 3.4). Notice that vertex 6 has no outneighbor and hence it is a *sink*. As a result the sixth row of the transition probability matrix P has entries all zero.

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

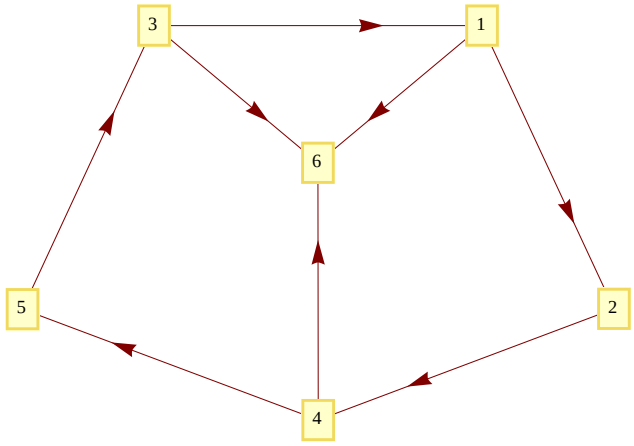


Figure 3.4: Directed graph with sink at vertex 6

Replacing row six of P by $\frac{1}{6}e^T$ yields the revised transition probability matrix $\bar{P}$:

$$\bar{P} = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

The revised irreducible matrix is then given by

$$\tilde{P} = \alpha\bar{P} + (1-\alpha)ee^T/n,$$

for an appropriate choice of α.

Remark 3.3 *Once we fix the issue of irreducibility of the transition probability matrix of a digraph, we can define the Laplacian of a nonstrongly connected digraph in terms of the new transition probability matrix. This is really a huge advantage as most digraphs in practice are not strongly connected.*

Conclusion: We gave an exposition of basic notations and concepts in graph theory. In particular, we discussed about graphs and the Laplacian operators

associated with them. Three most useful matrices associated with graphs: the adjacency matrix, combinatorial Laplacian and the normalized Laplacian for both weighted and unweighted graphs were discussed. We also reviewed the spectral properties of the Laplacian operators, known Laplacian eigenvalue bounds, spectral decomposition theorem, and the variational principle. Infinite graphs and an essential self-adjointness of an infinite Laplacian are also discussed.

CHAPTER 4

POINTWISE SAMPLING ON COMBINATORIAL GRAPHS

Spectral graph theory has historically focussed on constructing, analyzing, and manipulating graphs as opposed to signals on graphs. It has proved particularly useful for the construction of expander graphs [94], graph visualization [95], spectral clustering [41, 42, 43], graph coloring, graph drawing [46, 47, 48, 49] and numerous other applications. In the area of signal processing on graphs, spectral graph theory has been leveraged as a tool to define frequency spectral expansion bases for graph Fourier transform. In the next section we will discuss the notion of Fourier transform in the graph setting and extend many of the important mathematical ideas and intuitions from the classical Fourier analysis.

Recently, a new approach to a sampling theory of Paley-Wiener functions on combinatorial graphs was developed in [22]. A notion of Paley-Winer spaces on combinatorial graphs was introduced. It was shown that functions from some of these spaces can be uniquely determined by their values on some sets of vertices called the *sampling* (or *uniqueness) sets*. Such sampling sets are described in terms of Poincaré-Wringer type inequalities. In particular, it was shown that every finite subset of a graph admits a Poincaré inequality. A reconstruction algorithm of Paley-Wiener functions from sampling sets has

been developed. The goal of this chapter is to extend the theory to a more general weighted combinatorial graphs. We will also develop a reconstruction algorithm to construct Paley-Wiener functions from their sampling sets. Our experiment also shows that the eigenfunctions of the graph Laplacian corresponding with the lower eigenvalues are smooth and oscillate slowly whereas the eigenfunctions corresponding with larger eigenvalues oscillate much more rapidly and cross zeros more often, confirming the interpretation of the graph Laplacian eigenvalues as notions of frequency. Our sampling technique is based on the graph Fourier transform, which will be discussed in the next section.

4.1 A graph Fourier transform and notion of frequency

The classical Fourier transform on $\mathbb{R}^n$ is still an area of active research, particularly concerning Fourier transformation on more general objects such as tempered distributions. For instance, by imposing some requirements on a distribution f, one can attempt to translate these requirements in terms of the Fourier transform of f. The Paley-Wiener theorem is a good example of this. It immediately implies that if f is a nonzero distribution of compact support (these include functions of compact support), then its Fourier transform is never compactly supported. This is a very elementary form of an uncertainty principle in a harmonic analysis setting.

For a continuous functions defined on the real line, the classical Fourier transform

$$\hat{f}(\xi) := \langle f, e^{2\pi i \xi t} \rangle = \int_{-\infty}^{\infty} f(t) e^{-i\xi t} \, dt \tag{4.1}$$

provides a way of expanding the function f on the whole real line $\mathbb{R}$ as superpositions of the basic functions (complex exponential) $e^{i\xi t}$, $\xi \in \mathbb{R}$. In other words, the complex exponential $e^{i\xi t}$ defining the Fourier transform are

eigenfunctions of the one-dimensional Laplacian operator :

$$-\Delta(e^{i\xi t}) = -\frac{\partial^2}{dt^2}e^{i\xi t} = \xi^2 e^{i\xi t}. \tag{4.2}$$

The inverse Fourier transform can thus be seen as the expansion of f in terms of the eigenfunctions of the Laplacian operator. In principle, the knowledge of the values $f(x)$ for all $x \in \mathbb{R}$ determines completely all the properties of f and also all properties of $\hat{f}$, because the Fourier transform is one-to-one. However, it is not always convenient, let alone easy, to extract properties of $\hat{f}$ by looking only at f. For example, it is very difficult to decide in terms of f alone, whether $\hat{f} \in L_p$ when $p \neq 2$.

Although it is generally difficult to recognize properties of f from properties of $\hat{f}$ and vice versa, they are governed by two basic principles: the smoothness and decay principle and the uncertainty principle. The former describes that if f is smooth, then $\hat{f}$ decays quickly; if f decays quickly, then $\hat{f}$ is smooth. The latter describes f and $\hat{f}$ cannot be simultaneously localized (cannot be simultaneously small).

The graph Fourier transform is defined in precise analogy. We can define the graph Fourier transform $\hat{f}$ of any function $f \in L_2(G)$ on the vertices of G as the expansion of f in terms of the eigenfunctions of the graph Laplacian operator. Since the Laplacian operator L is a real symmetric matrix, it has a complete set of orthonormal eigenfunctions $\{\phi_\ell\}_{\ell=1,2,\ldots,n}$.

Definition 4.1 *Let $G = (V, E, w)$ be a connected weighted graph with $|V| = n$. The graph Fourier transform of $f \in L_2(G)$ is defined as*

$$\widehat{f}(\lambda_\ell) = \langle f, \phi_\ell \rangle = \sum_{v \in V} f(v)\phi_\ell^*(v), \quad \ell = 1, 2, \ldots, n \tag{4.3}$$

where $\phi_\ell \in \mathbb{R}^n$ is the ℓ^{th} orthonormal eigenfunction corresponding to the eigenvalue λ_ℓ. The original function f (i.e, the inverse graph Fourier transform) is then given by

$$f(v) = \sum_{\ell=1}^{n} \langle f, \phi_\ell \rangle \phi_\ell(v) = \sum_{\ell=1}^{n} \widehat{f}(\lambda_\ell)\phi_\ell(v), \quad v \in V(G). \tag{4.4}$$

In classical Fourier analysis, the eigenvalues $\{\xi^2\}_{\xi\in\mathbb{R}}$ in eq.(4.2) carry a specific notion of frequency: for ξ close to zero (low frequencies), the associated complex exponential eigenfunctions are smooth, slowly oscillating functions, whereas for ξ far from zero (high frequencies), the associated complex exponential eigenfunctions oscillate much more rapidly. In the graph setting, the graph Laplacian eigenvalues and eigenfunctions provide a similar notion of frequency. For example, for connected unweighted graphs, the Laplacian eigenfunction ϕ_1 associated with the eigenvalue 0 is constant and equal to $\frac{1}{\sqrt{n}}$ at each vertex, i.e., $\phi_1 = \frac{1}{\sqrt{n}}.\mathbf{1}$. More generally, the graph Laplacian eigenfunctions associated with low frequencies (small eigenvalues)λ_ℓ vary slowly across the graph. If two vertices are connected by an edge with a large weight, the values of the eigenfunction at those locations are likely to be similar. The eigenvectors associated with larger eigenvalues oscillate more rapidly and are more likely to have dissimilar values on vertices connected by an edge with high weights. This is demonstrated in figures (4.1) and (4.2).

As we can see from these figures, the Laplacian eigenfunctions associated with larger eigenvalues cross zero more often, confirming the interpretation of the graph Laplacian eigenvalues as notions of frequency. The set of zero crossings of a signal f on a graph G is defined as

$$\mathcal{Z}_{\mathcal{G}}(f) := \{e = uv \in E(G) : f(u)f(v) < 0\};$$

that is, the set of edges connecting a vertex with a positive signal to a vertex with a negative signal.

The graph Fourier transform and its inverse give us a way to equivalently represent a signal in two different domains: the vertex domain and and the graph spectral domain. Analogous to the classical case, the graph Fourier coefficients of a smooth signal decay rapidly. Such signals are *compressible* as they can be closely approximated by just a few graph Fourier coefficients.

We recall from eq.(3.11) that the quadratic form of the Laplacian at $f \in$

$L_2(G)$ is given by

$$\langle f, Lf \rangle = \sum_{i \sim j} (f(i) - f(j))^2.$$

We denote the square root of the quadratic form by $||f||_G$. i.e.,

$$||f||_G = \langle f, Lf \rangle^{1/2}$$

and call it the 2-norm graph total variation.

The following theorem describes a bound for the Fourier transform of a signal.

Theorem 4.1 *For any $f \in L_2(G)$,*

$$|\widehat{f}(\lambda_i)| \leq \frac{||f||_G}{\sqrt{\lambda_i}} \tag{4.5}$$

Proof:

$$\lambda_i |\widehat{f}(\lambda_i)|^2 \leq \sum_{i=1}^{n} \lambda_i |\widehat{f}(\lambda_i)|^2 = f^T (\sum_{i=1}^{n} \lambda_i \phi_i \phi_i^T) f = f^T L f = ||f||_G^2$$

■

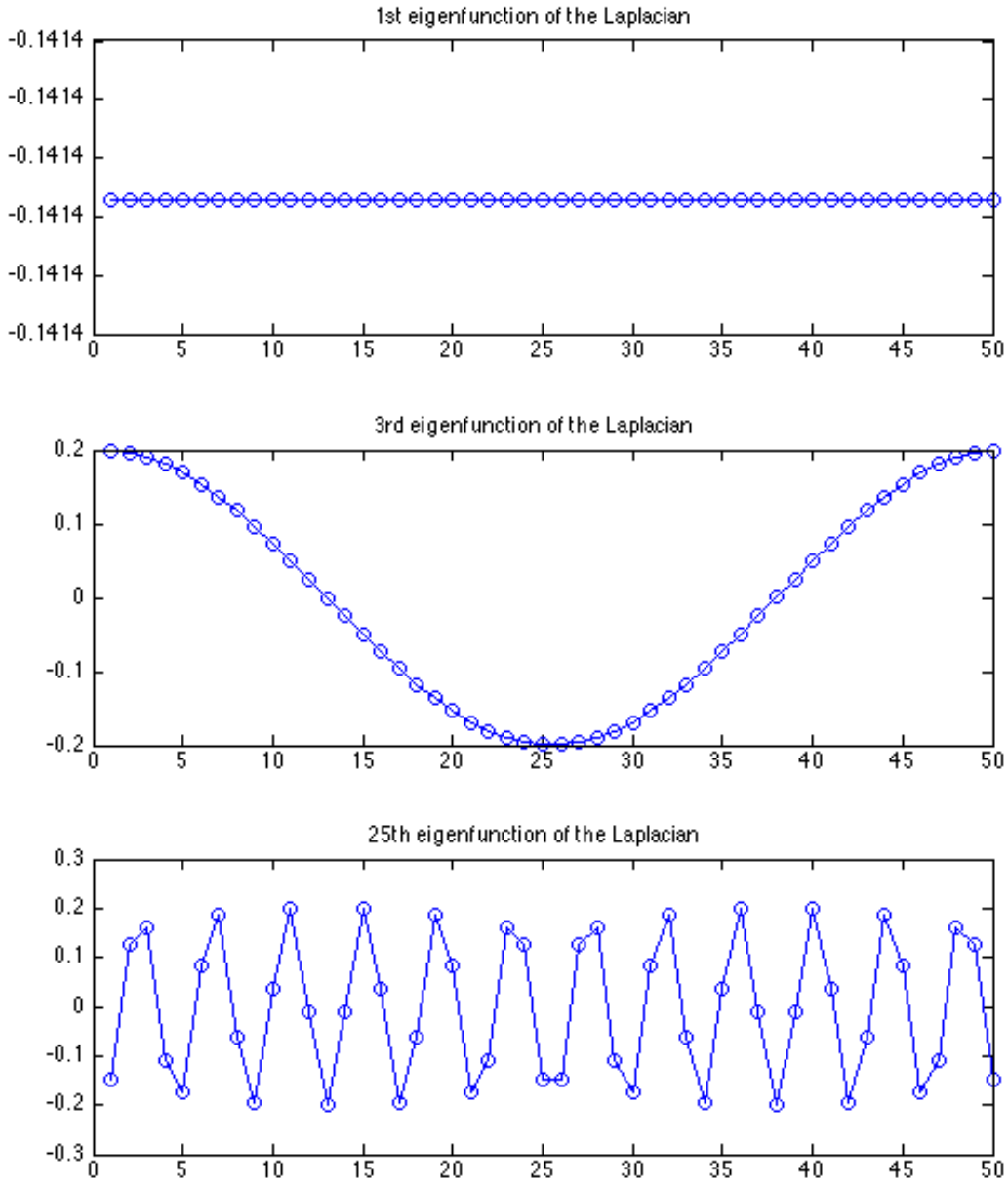


Figure 4.1: Some eigenfunctions of the Laplacian on a path graph

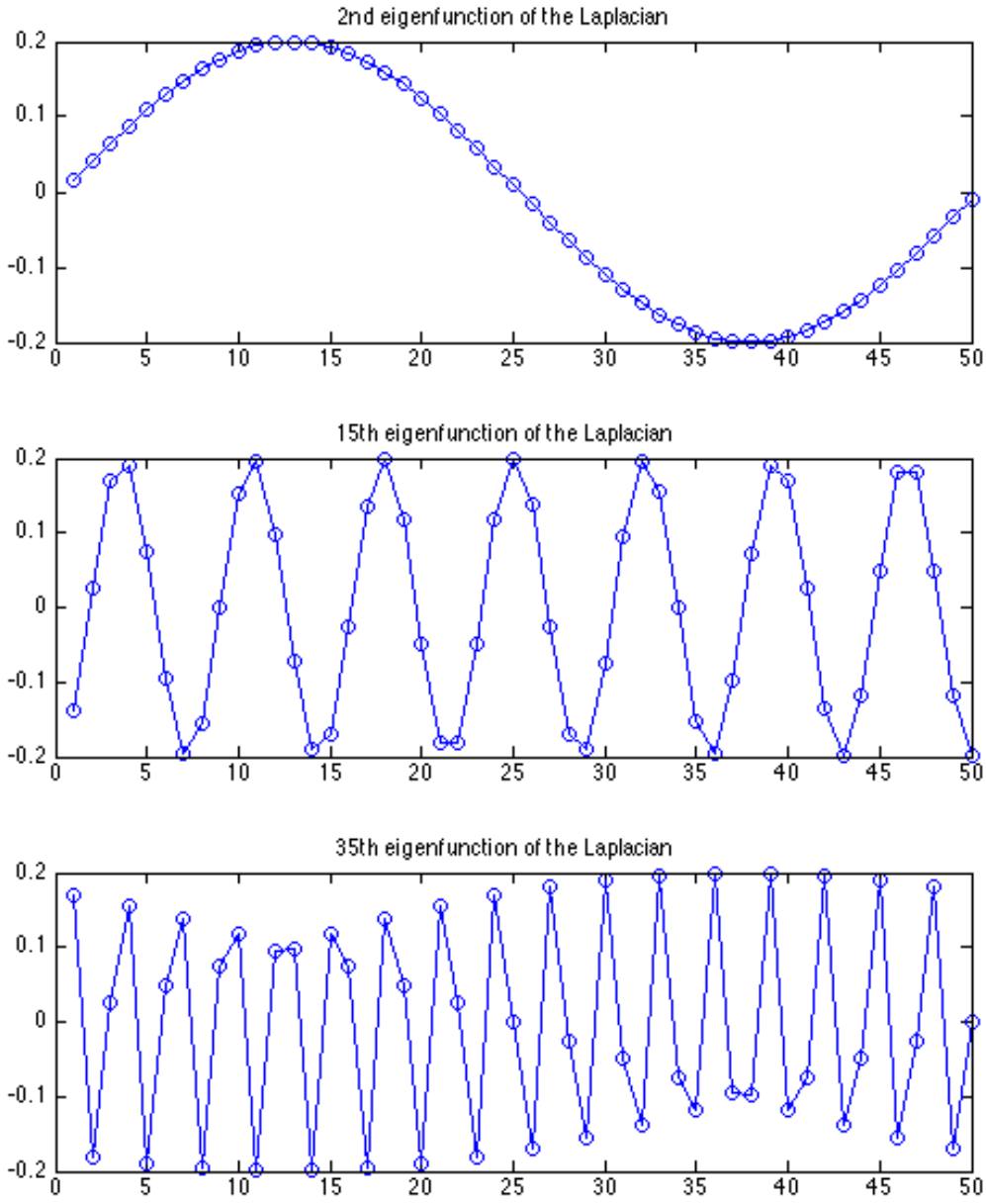


Figure 4.2: Some eigenfunctions of the Laplacian on a cycle graph

4.2 Bandlimited signals and sampling theory on graphs

The theory of bandlimited functions has been extensively studied for decades in signal analysis. The term is used to mean that the frequency content of a signal $f(t)$ is limited by certain bounds. More precisely, if $f(t)$ is a function of time, its Fourier transform

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{i\omega t}\, dt \tag{4.6}$$

is called the amplitude spectrum of f. It represents the frequency content of the signal. A signal is ω-bandlimited if $\hat{f}$ vanishes outside $[-\omega, \omega]$. The space of all ω-bandlimited signals is known as the **Paley-Wiener space** or **Bernstein space**. It is a subspace of $L_2(\mathbb{R})$ consisting of all functions whose Fourier transform is supported on the interval $[-\omega, \omega]$ and denoted by $PW_\omega(\mathbb{R})$.

The classical Shannon Sampling Theorem states that if a function f belongs to the space of bandlimited functions $PW_\omega(\mathbb{R})$, then it can be completely recovered from its uniform samples via the formula

$$f(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2\omega}\right) \frac{\sin(2\pi\omega t - \pi k)}{2\pi\omega t - \pi k} := \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2\omega}\right) sinc(2\pi\omega t - \pi k). \tag{4.7}$$

As mentioned in the introduction, the notion of Paley-Wiener spaces was recently extended to the space of functions defined on graphs [22]. The theory was developed using the fact that the graph Laplacian operator is a self-adjoint positive definite operator. A new characterization of the classical Paley-Wiener space that connects Paley-Wiener functions to analytic functions of a Cauchy problem involving Schrödinger equation was derived and a number of other properties of the space were discovered. We will discuss some of the fundamental properties of Paley-Wiener functions here just for the sake of completeness and refer the readers to [22] and [33] for details.

Definition 4.2 *Given $\omega \geq 0$, we will say that a signal f in $L_2(G)$ belongs to the Paley-Wiener space $PW_\omega(G)$ if the graph Fourier transform $\hat{f}$ has support*

in $[0, \omega]$. *In this case we say the function* f *is bandlimitted (specifically,* ω*-bandlimitted) or* f *is a Paley-Wiener function.*

Since the Laplacian $\mathcal{L}$ is a bounded operator, every function from $L_2(G)$ belongs to a certain Paley-Wiener space $PW_\omega(G)$ for some $\omega \in \sigma(\mathcal{L})$ and the following stratification holds:

$$L_2(G) = PW_{\omega_{\max}}(G) = \bigcup_{\sigma \in \mathcal{L}} PW_\omega(G), PW_{\omega_1}(G) \subseteq PW_{\omega_2}(G), \omega_1 < \omega_2,$$

where $\omega_{\max} = \sup_{\omega \in \sigma(\mathcal{L})} \omega$. The set $PW_\omega(G)$ is a linear closed subspace in $L_2(G)$.

Remark 4.1 *For a finite graph G, the Paley-Wiener space* $PW_\omega(G)$ *is a span of eigenfunctions of the Laplacian of G whose eigenvalues are less than or equal to* ω *and it is always nontrivial for every* $\omega \geq 0$.

Let $S \subset V(G)$. The space of all functions from $L_2(G)$ with support in S is denoted by:

$$L_2(S) = \{f \in L_2(G), f(v) = 0,\ v \in V(G) \backslash S\}. \tag{4.8}$$

Definition 4.3 *A set of vertices* $U \subset V(G)$ *is called a* uniqueness (sampling) *set for a space* $PW_\omega(G), \omega > 0$, *if for any two functions from* $PW_\omega(G)$, *the fact that they coincide on U implies that they coincide on V(G).*

Definition 4.4 *A set of vertices* $S \subset V(G)$ *is called a* Λ*-set if for any* $f \in L_2(S)$ *it admits a Poincaré inequality with a constant* $\Lambda > 0$ *such that*

$$||f|| \leq \Lambda ||\mathcal{L} f||, f \in L_2(S). \tag{4.9}$$

The infimum of all $\Lambda > 0$ for which S is is a Λ-set is called the *Poincaré constant* of the set S and denoted by $\Lambda(S)$.

Theorem 4.2 *Let* $G = (V, E, w)$ *be a weighted graph,* $\omega > 0$, *and* $\mathcal{L}$ *is the Laplacian of G. The following statements hold true:*

1. *$f \in PW_\omega(G)$ if and only if for all $t \in \mathbb{R}^+$ the following Bernstein inequality is satisfied*

$$||\mathcal{L}^t f|| \leq \omega^t ||f||, \quad f \in L_2(G). \tag{4.10}$$

2. *the norm of the operator $\mathcal{L}$ in the space $PW_\omega(G)$ is exactly ω.*

3. *$f \in PW_\omega(G)$ if and only if the following holds true:*

$$\lim_{t\to\infty} ||\mathcal{L}^t f|| = \omega, t \in \mathbb{R}^+, f \in L_2(G). \tag{4.11}$$

Λ-sets play an important role in sampling and approximation of graph signals. The following result partly explains the role of Λ-set in sampling and approximation.

Theorem 4.3 *Suppose $G = (V, E, w)$ is a connected weighted graph. If a set $S \subset V(G)$ is a Λ-set, then $U = V(G) \setminus S$ is a uniqueness set for any space $PW_\omega(G)$ with $\omega < \frac{1}{\Lambda}$.*

Proof: We must show that for any two arbitrary functions f and g in $PW_\omega(G)$,

$$f = g \text{ in } U \Longrightarrow f \equiv g \text{ in } V(G).$$

If $f, g \in PW_\omega(G)$, then $f - g \in PW_\omega(G)$. So by Theorem 4.2, we have

$$||\mathcal{L}(f - g)|| \leq \omega ||f - g||. \tag{4.12}$$

On the other hand, if f and g coincide in $U = V(G) \setminus S$, then $f - g$ belongs to $L_2(S)$. Since S is a Λ-set, it satisfies the Poincaré-inequality

$$||f - g|| \leq \Lambda ||\mathcal{L}(f - g)||.$$

Thus, if $f - g$ is not identically zero and $\omega < \frac{1}{\Lambda}$, we have the following inequalities:

$$||f-g|| \leq \Lambda||\mathcal{L}(f-g)|| \leq \Lambda\omega||f-g|| < ||f-g||, \tag{4.13}$$

which contradicts the assumption that $f-g$ is not identically zero. It completes the proof. ■

The following example provides a Poincaré-inequality on singleton sets and the next theorem generalizes this example in a more general setting.

Example : Let $G=(V,E,w)$ be a connected weighted graph and let $v \in V$ be any vertex and $S=\{v\}$. For any $\varphi \in L_2(S)$ we have

1. $||\varphi||_{L_2(G)} = \left(\sum_{u\in U} |\varphi(u)|^2\right)^{\frac{1}{2}} = (|\varphi(v)^2|)^{\frac{1}{2}} = |\varphi(v)|$
2. $\mathcal{L}\varphi(v) = \frac{1}{\sqrt{d(v)}} \sum_{u\sim v} \left(\frac{\varphi(v)}{\sqrt{d(v)}} - \frac{\varphi(u)}{\sqrt{d(u)}}\right) = \frac{1}{\sqrt{d(v)}} N(v) \frac{\varphi(v)}{\sqrt{d(v)}} = \frac{N(v)\varphi(v)}{d(v)}$, where $N(v)$ and $d(v)$ are the number of neighbors and weighted degree of vertex v respectively.
3. $\mathcal{L}\varphi(u) = \frac{1}{\sqrt{(u)}} \sum_{u\sim v} \frac{-\varphi(v)}{\sqrt{d(v)}} = \frac{-\varphi(v)}{\sqrt{d(u)\,d(v)}}$, for $u \sim v$, and $\mathcal{L}\varphi(u) = 0$ for all $u \neq v, u \nsim v$(u is not adjacent to v).

So we have the following:

$$\mathcal{L}\varphi(u) = \begin{cases} \frac{N(v)\varphi(v)}{d(v)} = \varrho(v)\varphi(v) & \text{if } u = v\ , \\ -\frac{\varphi(v)}{\sqrt{d(v)d(u)}} & \text{if } u \sim v, \\ 0 & \text{otherwise.} \end{cases} \tag{4.14}$$

Thus,

$$\begin{aligned}
||\mathcal{L}\varphi||_{L_2(G)} &= \left(\sum_{u\in V(G)} |\mathcal{L}\varphi(u)|^2\right)^{\frac{1}{2}} \\
&= \left(\frac{N^2(v)\varphi^2(v)}{d^2(v)} + \frac{\varphi^2(v)}{d(v)}\sum_{u\sim v}\frac{1}{d(u)}\right)^{\frac{1}{2}} \\
&= \left(\varrho^2(v)\varphi^2(v) + \frac{\varphi^2(v)}{d(v)}\sum_{u\sim v}\frac{1}{d(u)}\right)^{\frac{1}{2}} \\
&= \left(\varrho^2(v) + \frac{1}{d(v)}\sum_{u\sim v}\frac{1}{d(u)}\right)^{\frac{1}{2}} ||\varphi||_{L_2(G)} \\
&\geq \varrho(G)||\varphi||_{L_2(G)},
\end{aligned}$$

where $\varrho(v) = \frac{N(v)}{d(v)}$ and $\varrho(G) = \min \varrho(v), v \in V(G)$. Therefore

$$||\varphi||_{L_2(G)} \leq \frac{1}{\varrho(G)}||\mathcal{L}\varphi||_{L_2(G)}. \tag{4.15}$$

■

Definition 4.5 *Let $G = (V, E, w)$ be a connected weighted graph and $S \subset V$. The* vertex boundary *of S denoted by bS is the set of all vertices v not in S but adjacent to some vertex in S, i.e.,*

$$bS = \{v \notin S : (u, v) \in E(G), \textit{for some } u \in S\}.$$

Note that bS is always nonempty for a connected graph G. For a finite set S consider the set $\bar{S} = S \cup bS$ as an induced subgraph of G. Let $\Gamma(S)$ be a graph constructed in the following way: Take two copies of the induced graph $\bar{S}$, and denote them as $\bar{S}_1$ and $\bar{S}_2$, and identify every vertex $v \in bS \subset \bar{S}_1$ with the same vertex $v \in bS \subset \bar{S}_2$. The following theorem is a more general form of the above example.

Theorem 4.4 *Suppose $G = (V, E, w)$ is a connected weighted graph and $S \subset V(G)$ (finite or infinite) has the property that for any $v \in S$ its closure $\bar{v} = v \cup bv$ does not contain other points of S. Then S is a Λ-set with $\Lambda = \frac{1}{\varrho(G)}$.*

Proof: Let $\varphi \in L_2(S)$. For each $v \in S$, $\bar{v} \cap S = \{v\}$.

$$||\varphi||_{L_2(G)} = \left(\sum_{u \in V(G)} |\varphi(u)|^2 \right)^{\frac{1}{2}} = \left(\sum_{u \in S} |\varphi(u)|^2 \right)^{\frac{1}{2}}$$

(i) For each $v \in S$,

$$\mathcal{L}\varphi(v) = \frac{1}{\sqrt{d(v)}} \sum_{u \sim v} \left(\frac{\varphi(v)}{\sqrt{d(v)}} - \frac{\varphi(u)}{\sqrt{d(u)}} \right) = \frac{1}{\sqrt{d(v)}} N(v) \frac{\varphi(v)}{\sqrt{d(v)}} = \frac{N(v)}{d(v)} \varphi(v).$$

(ii) For each $v \in V(G) \backslash S$,

$$\mathcal{L}\varphi(v) = \frac{1}{\sqrt{d(v)}} \sum_{u \sim v} \left(\frac{\varphi(v)}{\sqrt{d(v)}} - \frac{\varphi(u)}{\sqrt{d(u)}} \right) = -\frac{1}{\sqrt{d(v)}} \sum_{u \sim v, u \in S} \frac{\varphi(u)}{\sqrt{d(u)}}.$$

Then we have,

$$\begin{aligned} ||\mathcal{L}\varphi||^2_{L_2(G)} = \left(\sum_{v \in V(G)} |\mathcal{L}\varphi(v)|^2 \right) & \\ &= \sum_{v \in S} |\mathcal{L}\varphi(v)|^2 + \sum_{u \in V(G) \backslash S} |\mathcal{L}\varphi(u)|^2 \\ &= \sum_{v \in S} \Big| \frac{N(v)}{d(v)} \varphi(v) \Big|^2 + \sum_{v \in V(G) \backslash S} \frac{1}{d(v)} \Big| \sum_{u \in S} \frac{\varphi(u)}{\sqrt{d(u)}} \Big|^2 \\ &\geq (\varrho(G))^2 ||\varphi||^2. \qquad \blacksquare \end{aligned}$$

Theorem 4.5 *Suppose $G = (V, E, w)$ is a connected weighted graph and $S \subset V(G)$ (finite or infinite) is a subset of V(G) satisfying the following three conditions:*

1. *every point from S is adjacent to a point from the boundary of S;*
2. *for every $v \in S$ there exists at least one adjacent point $v^* \in bS$ whose adjacency set intersects S only at v;*
3. *the number*

$$\eta_S = \sup_{v \in S} d(v^*)$$

is bounded.

Then S is a Λ-set with $\Lambda = \sqrt{\frac{\eta_S}{\varrho(G)}}$.

Proof: Let

$$S^* = \{u \in bS \,|\, \exists v \in S \text{ with } v^* = u\}.$$

By assumption of the theorem, S is non-empty. If $\varphi \in L_2(S)$, $v \in S$, we have

$$\mathcal{L}\varphi(v^*) = \frac{1}{\sqrt{d(v^*)}} \sum_{u \sim v^*} \left(\frac{\varphi(v^*)}{\sqrt{d(v^*)}} - \frac{\varphi(u)}{\sqrt{d(u)}} \right) = -\frac{\varphi(v)}{\sqrt{d(v^*)d(v)}},$$

and

$$\begin{aligned}
||\mathcal{L}\varphi||^2_{L_2(G)} &= \sum_{u \in V(G)} |\mathcal{L}\varphi(u)|^2 = \sum_{u \in S} |\mathcal{L}\varphi(u)|^2 + \sum_{u \in V(G)\setminus S} |\mathcal{L}\varphi(u)|^2 \\
&\geq \sum_{u \in V(G)\setminus S} |\mathcal{L}\varphi(u)|^2 \geq \sum_{v \in S} |\mathcal{L}\varphi(v^*)|^2 N(v) \\
&= \sum_{v \in S} \frac{|\varphi(v)|^2}{d(v)d(v^*)} N(v) = \sum_{v \in S} \frac{|\varphi(v)|^2}{d(v^*)} \frac{N(v)}{d(v)} \\
&\geq \left(\sum_{v \in S} \frac{|\varphi(v)|^2}{d(v^*)} \right) \varrho(G) \geq \frac{\varrho(G)}{\eta_S} \sum_{v \in S} |\varphi(v)|^2 \\
&= \frac{\varrho(G)}{\eta_S} \sum_{v \in V(G)} |\varphi(v)|^2 \\
&= \frac{\varrho(G)}{\eta_S} ||\varphi||^2.
\end{aligned}$$

Thus,

$$||\varphi|| \leq \sqrt{\frac{\eta_S}{\varrho(G)}} ||\mathcal{L}\varphi||. \qquad \blacksquare$$

Note that for unweighted graph, $\varrho(G) = 1$ and $\eta_S \geq 1$ and hence $||\varphi|| \leq \eta_S ||\mathcal{L}\varphi||$.

The following theorem allows to construct infinite Λ-sets from the finite ones.

Theorem 4.6 *Suppose that $G = (V, E, w)$ is a weighted graph and $\{S_j\}$ is a finite or an infinite sequence of disjoint subsets of vertices $S_j \subset V$ such that the sets $S_j \cup bS_j$ are pairwise disjoint. If each S_j is a Λ_j-set, then their union $S = \cup_j S_j$ is a Λ-set with $\Lambda = \sup_j \Lambda_j$.*

Theorem 4.7 *Suppose $G = (V, E, w)$ is a connected weighted graph and S is a nonempty finite subset of V(G). Then S is a Λ-set with $\Lambda = \frac{\sqrt{2}}{\lambda_1(\Gamma(S))}$.*

That is, for any $\varphi \in L_2(S)$, the inequality

$$||\varphi||_{L_2(G)} \leq \frac{\sqrt{2}}{\lambda_1(\Gamma(S))}||\mathcal{L}\varphi||_{L_2(G)}$$

holds true.

Proof: Construct an embedding of the space $L_2(S)$ into the space $L_2(\Gamma(S))$ as follows: If $\varphi \in L_2(S)$, then its image $F_\varphi \in L_2(\Gamma(S))$ is defined by

$$F_\varphi(v) = \begin{cases} \varphi(v) & \text{if } v \in \bar{S}_1, \\ -\varphi(v) & \text{if } v \in \bar{S}_2 \end{cases} \tag{4.16}$$

Then we have

$$\begin{aligned} (i) \quad ||F_\varphi||^2_{L_2(\Gamma(S))} &= \sum_{v\in\Gamma(S)} |F_\varphi(v)|^2 \\ &= \sum_{v\in\bar{S}_1} |F_\varphi(v)|^2 + \sum_{v\in S_2} |F_\varphi(v)|^2 \\ &= \sum_{v\in S_1} |\varphi(v)|^2 + \sum_{v\in S_2} |-\varphi(v)|^2 \\ &= 2\sum_{v\in S} |\varphi(v)|^2 \\ &= 2\sum_{v\in V(G)} |\varphi(v)|^2 \\ &= 2||\varphi||^2. \end{aligned}$$

Thus,

$$||F_\varphi||_{L_2(\Gamma(S))} = \sqrt{2}||\varphi||_{L_2(G)}. \tag{4.17}$$

(ii) Let $N_{\Gamma(S)}(v)$, $N_{\bar{S}}(v)$, $N_G(v)$ be the combinatorial degrees (or number of neighbors) of v in $\Gamma(S)$, $\bar{S}$ and G respectively. It is easy to see that $N_{\Gamma(S)}(v) \leq 2N_G(v)$. As a result

$$||\mathcal{L}_{\Gamma(S)}F_\varphi||_{L_2}(\Gamma(S)) \leq 2||\mathcal{L}_G\varphi||_{L_2(G)}. \tag{4.18}$$

(iii) We know that the vector $\left(\sqrt{d(v_1)}, \ldots, \sqrt{d(v_n)}\right)$ is an eigenvector for the weighted Laplacian $\mathcal{L}$ corresponding to the eigenvalue zero. Thus

$$\Psi_0(v) = \sqrt{d_{\Gamma(S)}(v)}, v \in \Gamma(S)$$

is the eigenfunction of $\mathcal{L}_{\Gamma(S)}$ which corresponds to the eigenvalue 0. Since every function F_φ is "odd" (in the sense that $\varphi(v) = -\varphi(v)$ when v is considered as an element in $\bar{S_2}$) it is orthogonal to the subspace spanned by Ψ_0. Now let $\{\Psi_k\}, k = 0, 1, ..., N$ is a complete orthonormal system of eigenfunctions of $\mathcal{L}_{\Gamma(S)}$. (Note: Any real symmetric matrix has an orthonormal basis of real eigenvectors with its associated eigenvalues.) Hence we have

$$F_\varphi = \sum_{j=1}^{N} \langle F_{\varphi,\Psi_j\rangle}\Psi_j;$$

$$\mathcal{L}_{\Gamma(S)}F_\varphi = \sum_{j=1}^{N} \lambda_j(\Gamma(S))\langle F_{\varphi,\Psi_j\rangle}\Psi_j.$$

Thus,

$$||\mathcal{L}_{\Gamma(S)}F_\varphi||^2{}_{L_2}(\Gamma(S)) = \sum_{j=1}^{N} {\lambda_j}^2(\Gamma(S))|\langle F_{\varphi,\Psi_j\rangle|}{}^2 \geq {\lambda_1}^2(\Gamma(S))||F_\varphi||^2{}_{L_2(\Gamma(S))}.$$

That is

$$\lambda_1(\Gamma(S))||F_\varphi||_{L_2(\Gamma(S))} \leq ||\mathcal{L}_{\Gamma(S)}F_\varphi||_{L_2}(\Gamma(S)). \tag{4.19}$$

Now combining equations (4.17) - (4.19) we obtain the desired result

$$||\varphi||_{L_2(G)} \leq \frac{\sqrt{2}}{\lambda_1(\Gamma(S))}||\mathcal{L}\varphi||_{L_2(G)}. \qquad \blacksquare$$

Let us now see some examples of Λ-sets.

Some Examples:

Here we will give some examples of Λ sets with an explicit value of Λ for infinite graphs. In the case of finite graphs, we will show a nice method of computing the optimal value of Λ in chapter 5.

Example 1. Consider a line graph G = $\mathbb{Z}$ whose vertices correspond to the set of integers $\mathbb{Z}$. For any finite subset S of successive vertices of the graph $\mathbb{Z}$ the following inequality holds true:

$$||\varphi|| \leq \frac{1}{2} \sin^{-2} \frac{\pi}{2|S|+2} ||\mathcal{L}\varphi||, \quad \varphi \in L_2(S).$$

In the following example, we will state explicit constants for the Poincaré inequality for some specific finite sets of points on homogeneous trees. Consider a homogeneous tree of order $q+1$. We will say that the roots of this tree belongs to the level zero, the next q vertices belong to the level one, the next q^2 belong to the level two and so on. A level of order m will be denoted as l_m.

Example 2. On a homogeneous tree G of order $q+1$ for any level $S = l_m$ of order m the following Poincaré inequality holds true:

$$||\varphi|| \leq \left(1 + \frac{q}{(q+1)^2}\right)^{-\frac{1}{2}} ||\mathcal{L}\varphi||, \quad \varphi \in L_2(S).$$

4.3 Sampling on Directed Graphs

The space of Paley-Wiener functions and a corresponding sampling theory for undirected graphs was discussed in sections 4.1 and 4.2. The notion was defined using the fact that the graph Laplacian operator is a self-adjoint semi-positive definite operator in the Hilbert space $L_2(G)$. In the present chapter we will extend the idea to directed graphs. Since the Laplacian operator on directed graphs is self-adjoint positive definite as in the case of undirected Laplacian, most of the notions and subsequent theories developed for undirected graphs carry over to directed graphs.

For instance, given an $\omega \geq 0$ we will say that a function $f \in L_2(G)$ belongs to the *Paley-Wiener Space* $PW_\omega(G)$ if its Fourier transform $\hat{f}$ has support in $[0, \omega]$. The notions of uniqueness sets, Λ-set, Poincaré inequalities, Plancherel-Polya inequalities described before will be defined in the same way. Moreover, when G is finite, $\sigma(\mathcal{L})$ is a discrete set and the Paley-Wiener Space $PW_\omega(G)$

is a span of eigenfunctions of the Laplacian whose eigenvalues are less than or equal to ω.

The main purpose of this section is to give some examples of Λ-sets on directed graphs.

Example 1: Consider a strongly connected k-regular digraph $G = (V, E)$ of order n and let v_0 be a vertex in V. Let $S = \{v_0\}$. We will show that S is a Λ-set with $\Lambda(S) = \Lambda = \frac{1}{\sqrt{1+\frac{1}{2k}}}$.

Proof: It is known that for strongly connected regular digraphs of order n and degree k the Perron vector ϕ and the transition probability p are given by $\phi(u) = \frac{1}{n}$ and $p(u, v) = \frac{1}{k}$ for each vertex $u, v \in V$. So for each $f \in L_2(S)$ we have

$$(i)\ (\mathcal{L}f)(v_0) = f(v_0) - \frac{1}{2}\left(\sum_{u:u\to v_0} \frac{\phi(u)\,p(u,v_0)\,f(u)}{\sqrt{\phi(u)\,\phi(v_0)}} + \sum_{w:v_0\to w} \frac{\phi(v_0)\,p(v_0,w)f(w)}{\sqrt{\phi(w)\,\phi(v_0)}}\right)$$
$$= f(v_0)$$

$$(ii).\ (\mathcal{L}f)(w) = f(w) - \frac{1}{2}\left(\sum_{u:u\to w} \frac{\phi(u)\,p(u,w)\,f(u)}{\sqrt{\phi(u)\,\phi(w)}} + \sum_{t:w\to t} \frac{\phi(w)\,p(w,t)f(t)}{\sqrt{\phi(w)\,\phi(t)}}\right)$$
$$= -\frac{f(v_0)}{2k}$$

for each w such that $w \to v_0$.

(iii). Similarly, for each u such that $v_0 \to u$ we have $(\mathcal{L}f)(u) = -\frac{f(v_0)}{2k}$,

(iv). Finally for each $u \neq v_0, u \nrightarrow v_0$ or $u \nleftarrow v_0$, $(\mathcal{L}f)(u) = 0$.

Thus,

$$||\mathcal{L}f||^2_{L_2(G)} = \sum_{u\in V} |\mathcal{L}f(u)|^2$$
$$= |f(v_0)|^2 + \sum_{w:w\to v_0} \frac{|f(v_0)|^2}{4k^2} + \sum_{u:v_0\to u} \frac{|f(v_0)|^2}{4k^2}$$
$$= |f(v_0)|^2 + k\left(\frac{|f(v_0)|^2}{4k^2}\right) + k\left(\frac{|f(v_0)|^2}{4k^2}\right)$$
$$= |f(v_0)|^2(1 + \frac{1}{2k}) = (1 + \frac{1}{2k})||f||^2_{L_2(G)},$$

which implies

$$||f|| \leq \frac{1}{\sqrt{1+\frac{1}{2k}}}||\mathcal{L}f||. \tag{4.20}$$

■

Example 2: Let v be any vertex in a strongly connected digraph G (not necessarily regular) and let $S = \{v\}$. It is an easy calculation to show that for any $f \in L_2(S)$,

$$\mathcal{L}f(u) = \begin{cases} f(v) & \text{if } u = v \ ; \\ -\frac{f(v)}{2\,d(u)}\sqrt{\frac{\phi(u)}{\phi(v)}} & \text{if } (u,v) \in E(G); \\ -\frac{f(v)}{2\,d(v)}\sqrt{\frac{\phi(v)}{\phi(u)}} & \text{if } (v,u) \in E(G); \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\Lambda(S) = \Lambda(v) = \frac{1}{\sqrt{1+\frac{1}{4}\left(\sum_{u:\ u\to v}\frac{\phi(u)}{\phi(v)\ d(u)^2}+\sum_{w:\ v\to w}\frac{\phi(v)}{\phi(u)\ d(v)^2}\right)}}. \tag{4.21}$$

Note that equation (4.20) can be recovered from (4.21) when G is regular.

Definition 4.6 *Given a digraph $G = (V, E)$, and any vertex $v \in V(G)$,*

- *the set of vertices $u \in V(G)$ such that $(u, v) \in E(G)$ is called the* in-neighbor *of v and denoted by $N^-(v)$.*
- *the set of vertices $u \in V(G)$ such that $(v, u) \in E(G)$ is called the* out-neighbor *of v and denoted by $N^+(v)$.*
- *we call a nonempty set $S \subset V(G)$* isolated *if for each $v \in S$ it contains neither the in-neighbor of v nor the out-neighbor of v.*

Suppose G is a digraph and $S \subset V(G)$. We will define the edge and vertex boundaries of S as follows.

(i) The *edge out-boundary* of S, denoted by ∂S, consists of all edges $(u, v) \in E(G)$ with $u \in S$ and $v \notin S$. i.e.,

$$\partial S = \{(u, v) \in E(G) \mid u \in S,\ v \notin S\}.$$

(ii) The *edge in-boundary* of S, denoted by $\partial \bar{S}$, is defined similarly as

$$\partial \bar{S} = \{(u, v) \in E(G) \mid u \in V \backslash S,\ v \in S\} = \{(u, v) \in E(G) | v \in S,\ u \notin S\}.$$

(iii) The *vertex out-boundary* of S, denoted by δS, is the set of all vertices $v \in V \setminus S$ such that $(u, v) \in \partial S$ for some $u \in S$. i.e.,

$$\delta S = \{v \in V \setminus S : (u, v) \in \partial S, \text{for some } u \in S\}.$$

(iv) The *vertex in-boundary* of S, denoted by $\delta \bar{S}$, is the set of all vertices $u \in V \setminus S$ such that $(u, v) \in \partial \bar{S}$ for some $v \in S$. i.e.,

$$\delta \bar{S} = \{u : (u, v) \in \partial \bar{S}\}.$$

For the sake of conciseness, in the following, unless otherwise stated, we refer to the out-boundary or in-boundary followed by the appropriate notations without specifying vertex or edge. Note, for directed graphs in-boundary and out-boundary can be quite different. The following theorem shows that every isolated set is a Λ-set. The theorem helps to construct *infinite* Λ-sets.

Theorem 4.8 *Suppose that $G = (V, E)$ is a strongly connected digraph and $S \subset V(G)$ is an isolated subset of V(G). Then S is a Λ-set with $\Lambda = 1$.*

Proof: For each $\varphi \in L_2(S)$ and for each $v \in S$ we have

$$(\mathcal{L}f)(v) = f(v) - \frac{1}{2}\left(\sum_{u:u\to v} \frac{\phi(u)\, p(u,v)\, f(u)}{\sqrt{\phi(u)\,\phi(v)}} + \sum_{w:v\to w} \frac{\phi(v)\, p(v,w) f(w)}{\sqrt{\phi(w)\,\phi(v)}}\right) = f(v)$$

since $f(u) = 0 = f(w)\ \forall u, w \in V$ with $u \to v$ and $v \to w$.
On the other hand, by the hypothesis, for each $v \in V \setminus S$, all its in-neighbors

and out-neighbors of v belong to $V \setminus S$. So $f(v) = f(u) = f(w) = 0 \; \forall u, w \in V$ with $u \to v$ and $v \to w$ and therefore

$$(\mathcal{L}f)(v) = f(v) - \frac{1}{2}\left(\sum_{u:u\to v} \frac{\phi(u)\, p(u,v)\, f(u)}{\sqrt{\phi(u)\,\phi(v)}} + \sum_{w:v\to w} \frac{\phi(v)\, p(v,w) f(w)}{\sqrt{\phi(w)\,\phi(v)}} \right) = 0.$$

Therefore,

$$\begin{aligned} ||\mathcal{L}f||^2 &= \sum_{u\in V} |\mathcal{L}f(u)|^2 \\ &= \sum_{u\in S} |\mathcal{L}f(u)|^2 + \sum_{u\in V\setminus S} |\mathcal{L}f(u)|^2 \\ &= \sum_{u\in S} |\mathcal{L}f(u)|^2 \\ &= \sum_{u\in S} |f(u)|^2 \\ &= ||f||^2, \end{aligned}$$

which proves that $||f|| \leq ||\mathcal{L}f||$. ■

Theorem 4.9 *Suppose that for a set of vertices $S \subset V(G)$ (finite or infinite) the following conditions hold true:*

1) *for every point $v \in S$ there is a point $u \in \delta S$ such that $(v, u) \in \partial S$. That is, every point in S has an out-neighbor on the boundary of S.*

2) *for every $v \in S$ there exists at least one $u_v \in \delta S$ such that $(v, u_v) \in \partial S$, $N^-(u_v) \cap S = \{v\}$, and $N^+(u_v) \cap S = \emptyset$. That is, every point $v \in S$ has at least one out-neighbor u_v on the boundary δS whose in-neighbor intersects S only over v and whose out-neighbor is disjoint from S.*

Then S is a Λ-set with $\Lambda = 2\,{\Lambda_S}^{-1}$, where

$$\Lambda_S = \inf_{v\in S} \frac{\phi(v)}{d(v)}.$$

Proof: By hypothesis there exist a subset $S^* \subset \delta S$ such that for every vertex $v \in S$ there exists at least one point $u_v \in S^*$ whose in-neighbor intersects S

only over v and whose out-neighbor is disjoint from S. Now for each $\varphi \in L_2(S)$, and $u_v \in S^*,\ v \in S$, we have $\varphi(u_v) = 0$ and

$$\mathcal{L}\varphi(u_v) = \varphi(u_v) - \frac{1}{2}\left(\sum_{w:w\to u_v} \frac{\phi(w)\, p(w,u_v)\, \varphi(w)}{\sqrt{\phi(u_v)\,\phi(w)}} + \underbrace{\sum_{z:u_v\to z} \frac{\phi(u_v)\, p(u_v,z)\varphi(z)}{\sqrt{\phi(u_v)\,\phi(z)}}}_{=0 \text{ since } N^+(u_v)\cap S=\emptyset} \right)$$
$$= -\frac{1}{2}\left(\frac{\phi(v)\, p(v,u_v)\, \varphi(v)}{\sqrt{\phi(u_v)\,\phi(v)}} \right).$$

Since (by the hypothesis) for every $v \in S$ there exists at least one vertex u_v in S^* such that $(v,u_v) \in \partial S$ and since $\phi(v) < 1$ for all $v \in V(G)$ we have

$$||\mathcal{L}\varphi|| = \left(\sum_{v\in V(G)} |\mathcal{L}\varphi(v)|^2 \right)^{1/2} \geq \left(\sum_{v\in S} |\mathcal{L}\varphi(u_v)|^2 \right)^{1/2}$$
$$\geq \left(\sum_{v\in S} |\mathcal{L}\varphi(u_v)|^2 \phi(u_v)\,\phi(v) \right)^{1/2}$$
$$\geq \Lambda_S \frac{1}{2} ||\varphi||.$$

■

Theorem 4.10 *Let $G = (V,E)$ is a strongly connected digraph. Suppose that $\{S_j\}$ is a finite sequence of disjoint subsets of vertices $S_j \subset V$ such that the sets $cl(S_j) = S_j \cup \delta(S_j) \cup \delta(\bar{S}_j)$ are pairwise disjoint.*
Then if a set S_j has type Λ_j, then their union $S = \bigcup_j S_j$ is a set of type $\Lambda = \sup_j \Lambda_j$.

Proof. Since the sets S_j are disjoint, every $\varphi \in L_2(S), S = \bigcup_j S_j$, is a sum of functions $\varphi_j \in L_2(S_j)$ which are pairwise orthogonal. Further since the sets $cl(S_j)$ are pairwise disjoint, one can easily verify that the functions $\mathcal{L}\varphi_j$ are also orthogonal. Therefore, we have

$$||\varphi||^2 = \sum_j ||\varphi_j||^2 \leq \sum_j \Lambda_j^2 ||\mathcal{L}\varphi_j||^2 \leq \Lambda^2 ||\mathcal{L}\varphi||^2, \quad \text{where } \Lambda = \sup_j \Lambda_j.$$

This completes the proof. ■

Conclusion: A graph Fourier transform which resembles in many ways to the classical Fourier transform is defined using the graph Laplacian eigenfunctions. It is explained here why the Laplacian eigenvalues and eigenfunctions provide a similar notion of frequency and we gave experimental results supporting the theory behind it. Bandlimited graph signals are defined and a sampling theory on combinatorial graphs (both directed and undirected) developed. It is shown that every subset of the vertex set is a Λ-set for some appropriate Λ, and a bound for Λ is given although we do not know if this bound is optimal at this point.

CHAPTER 5

MULTIRESOLUTION ON WEIGHTED GRAPHS

Graphs are natural ways to represent data in many domains. For example, consider a data set with N elements, for which some information about the relationship between the data elements is known. This relational information can be represented by a graph $G = (V, E)$, where $V = V(G) = \{v_0, \ldots, v_{N-1}\}$ is the set of vertices and $E = E(G)$ is the set of edges or links connecting these vertices. Each dataset element corresponds to node v_n. Since data elements can be related to each other differently, in general, G is a *weighted* graph. The weight of the edge connecting two nodes u and v is denoted by $w(u, v)$. The degree $\mu(v)$ of the vertex v is the sum of the edge weights incident to node v. The adjacency matrix W of the graph is an $N \times N$ matrix such that $W(u, v) = w(u, v)$. The data on the graphs is often represented as a scalar or vector valued function attached to the vertices of the graph. The set of all complex valued functions f on $V(G)$ is denoted by $L_2(G)$.

The space $L_2(G)$ is the Hilbert space of all complex-valued functions with the following inner product

$$\langle f, g \rangle_{L_2(G)} = \langle f, g \rangle = \sum_{v \in V(G)} f(v)\overline{g(v)}\,\mu(v). \tag{5.1}$$

The weighted Laplacian operator $\mathcal{L}$ is defined on $L_2(G)$ via

$$(\mathcal{L}f)(v) = \sum_{u \in V(G)} (f(v) - f(u))w(v,u) \ . \tag{5.2}$$

The Laplacian is a well-studied object; it is known to be a positive-semidefinite self-adjoint *bounded* operator. Thus, it has N real and nonnegative eigenvalues. Moreover, since $\mathcal{L}\mathbf{1} = 0$, where $\mathbf{1} = (1, 1, \ldots, 1)$, is the all 1 constant function, zero is an eigenvalue of $\mathcal{L}$ corresponding to the eigenfunction $\mathbf{1}$.

According to [18] if for an *infinite* graph there exists a $C > 0$ such that the degrees are uniformly bounded

$$\mu(u) = \sum_{u \in V(G)} w(u,v) \leq C \qquad \forall u \in V(G), \tag{5.3}$$

then the operator which is defined by eq.(5.2) on functions with compact supports has a unique positive-semidefinite self-adjoint *bounded* extension $\mathcal{L}$ which is acting according to (5.2). Now let $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}$ be the set of eigenvalues of $\mathcal{L}$ and let $e_{\lambda_0}, ..., e_{\lambda_{N-1}}$ be an orthonormal complete set of eigenfunctions.

In classical signal analysis, a bandlimited function (signal) f on $\mathbb{R}$ may be explicitly reconstructed from its values $\{f(k\delta)|k \in \mathbb{Z}\}$ on an appropriately spaced lattice ($\delta = \frac{1}{2B}$ if B is the bandlimit) in terms of the sinc function $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$. This is the classical Shannon's sampling theorem [82]. We are interested in the discrete analog of Shannon's theorem, where we replace $\mathbb{R}$ by a graph G and the classical Laplace operator by the graph Laplacian operator $\mathcal{L}$. A signal is said to be ω-*bandlimited* if its Fourier transform (GFT) has support only on $[0, \omega]$. In other words, f is bandlimited means that the expansion of f as an infinite series in terms of eigenfunctions $\{e_k\}$ of $\mathcal{L}$ with $\mathcal{L}e_k = \lambda_k e_k$ is actually a finite sum where the eigenvalues λ_k satisfy $\lambda_k \leq \omega$, with ω being the bandlimit. The space of ω-bandlimited signals is called the **Paley-Wiener space** and is denoted by $PW_\omega(G)$.

5.1 Uniqueness (sampling) sets

We have seen in chapter 4 that a subset of vertices $U \subset V(G)$ is a uniqueness set (also known as a sampling set) for a space $PW_\omega(G),\ \omega > 0$, if for any two signals from $PW_\omega(G)$, the fact that they coincide on U implies that they coincide on $V(G)$. This definition implies that it is sufficient to know the values of a ω-bandlimited signal only on the uniqueness set U. However, in practice, given a $\omega > 0$ (or given a ω-bandlimited signal), determining a uniqueness set for space $PW_\omega(G)$ (or a sampling set for f) is a difficult task. In fact, in graph signal approximation, values of signals are usually known on some subset of the vertex set and the real problem is to determine which signals can be well approximated just based on the knowledge of their values on the given subset. In other words, for what values of ω does this subset be a uniqueness set for $PW_\omega(G)$? Obviously, it is not possible to perfectly reconstruct a random signal from its values on any proper subset of the graph. In fact, in [22], it was shown that signals which involve only low frequencies can be perfectly reconstructed from their values on some subsets of the vertices. The next theorem shows that every subset U of the vertex set $V(G)$ is a uniqueness set. Thus, the problem of signal interpolation can be posed as the problem of first defining the set of nodes with known sample values as a uniqueness set U, then identifying the maximum ω such that U is a uniqueness set for $PW_\omega(G)$, and then reconstructing the signal values on the complement set $V(G)\backslash U$.

Definition 5.1 *(Λ-**Set**) A subset of nodes $S \subset V(G)$ is a Λ-set if any $\varphi \in L_2(S)$ admits a Poincaré inequality with a constant $\Lambda > 0$, i.e.,*

$$\|\varphi\| \leq \Lambda \|\mathcal{L}\varphi\|, \quad \varphi \in L_2(S). \tag{5.4}$$

Theorem 5.1 *If $\omega > 0$ and $S \subset V(G)$ is a Λ-set with $\omega < 1/\Lambda$, then $U = V(G)\backslash S$ is a uniqueness set for $PW_\omega(G)$.*

The following Lemma will be used below.

Lemma 5.1 *[23] If A is a bounded self-adjoint positive definite operator in a Hilbert space H and for any $\varphi \in H$ and a constant $a > 0$ the following inequality holds true:*

$$\|\varphi\| \leq a\|A\varphi\|,$$

then for the same $\varphi \in H$, and all $k = 2^l, l = 0, 1, 2, ...$ the following inequality holds:

$$\|\varphi\| \leq a^k\|A^k\varphi\|.$$

Theorem 5.2 *Let $U \subset V(G)$ and $S = V(G)\backslash U$. Let $\mathcal{M}$ be a matrix which is obtained from the matrix of $\mathcal{L}$ by replacing by zero columns and rows corresponding to the set U.*

Then U is a uniqueness set for all signals $f \in PW_\omega(G)$ with $\omega < \sigma$ where σ is the smallest positive eigenvalue of $\mathcal{M}$.

Proof.

According to Definition 5.1 and Theorem 5.1, it suffices to show that S is a Λ-set for $\Lambda = \frac{1}{\sigma}$, i.e., for any $\varphi \in L_2(G)$ whose restriction to U is zero one has

$$\|\varphi\| \leq \frac{1}{\sigma}\|\mathcal{L}\varphi\|, \quad \varphi|_U = 0. \tag{5.5}$$

Note that, since matrix of $\mathcal{L}$ is symmetric and since in the latter we replace by zero columns and rows with the same set of indices the matrix $\mathcal{M}$ is also symmetric. Since $\mathcal{L}$ is non-negative the matrix $\mathcal{M}$ is also non-negative. It follows from the fact that if φ belongs to the subspace $L_2(S)$ of all $\varphi \in L_2(G)$ such that $\varphi|_U = 0$, then

$$\langle \mathcal{M}\varphi, \varphi\rangle = \langle \mathcal{L}\varphi, \varphi\rangle \geq 0.$$

Remark 5.1 *Warning: the equality*

$$\langle \mathcal{M}\varphi, \varphi\rangle = \langle \mathcal{L}\varphi, \varphi\rangle$$

holds for every φ in the subspace $L_2(S)$, but in general even for functions in $L_2(S)$ there is no equality

$$\mathcal{M}\varphi = \mathcal{L}\varphi.$$

Now we are going to show that $\mathcal{M}_S = \mathcal{M}|_S$ is strictly positive on $L_2(S)$. Indeed, it is clear that the subspace $L_2(S)$ of all $\varphi \in L_2(G)$ such that $\varphi|_U = 0$ is invariant with respect to $\mathcal{M}$. This fact allows to identify $\mathcal{M}$ with an operator $\mathcal{M}_S$ in $L_2(S)$.

If $\varphi \in L_2(S)$ is not identically zero and $\mathcal{M}\varphi = 0$, then

$$0 = \langle \mathcal{M}\varphi, \varphi \rangle = \langle \mathcal{L}\varphi, \varphi \rangle = \left\langle \mathcal{L}^{1/2}\varphi, \mathcal{L}^{1/2}\varphi \right\rangle = \|\mathcal{L}^{1/2}\varphi\|^2,$$

which implies that $\mathcal{L}^{1/2}\varphi = 0$ and then $\mathcal{L}\varphi = 0$. As the formula

$$\mathcal{L}\varphi(v) = \sum_{u \sim v} w(u,v)\left(\varphi(v) - \varphi(u)\right)$$

shows only functions which are constant on the entire graph belong to the kernel of $\mathcal{L}$. Since constants do not belong to $L_2(S)$ it implies strict positivity of the operator $\mathcal{M}_S$ on $L_2(S)$.

Let $0 < \sigma = \sigma_0 \leq \sigma_1 \leq ... \leq \sigma_{m-1}$, $m = |S|$, be the set of eigenvalues of $\mathcal{M}_S$ counting with their multiplicities and $e_0, ..., e_{m-1}$ be the corresponding set of orthonormal eigenvectors that forms a basis in $L_2(S)$.

Since $\mathcal{L}$ is symmetric and non-negative, it has a well defined positive square root $\mathcal{L}^{1/2}$. For $\varphi \in L_2(S)$ we have

$$\frac{\|\mathcal{L}^{1/2}\varphi\|^2}{\|\varphi\|^2} = \frac{\left\langle \mathcal{L}^{1/2}\varphi, \mathcal{L}^{1/2}\varphi \right\rangle}{\|\varphi\|^2} = \frac{\langle \mathcal{L}\varphi, \varphi \rangle}{\|\varphi\|^2} = \frac{\langle \mathcal{M}_S\varphi, \varphi \rangle}{\|\varphi\|^2}. \tag{5.6}$$

If $\varphi = \sum_{j=0}^{m-1} c_j e_j$ where $c_j = \langle \varphi, e_j \rangle$, then

$$\mathcal{M}_S\varphi = \sum_{j=0}^{m-1} \sigma_j c_j e_j,$$

and by Parseval equality

$$\langle \mathcal{M}_S\varphi, \varphi \rangle = \sum_{j=0}^{m-1} \sigma_j |c_j|^2 \geq \sigma \|\varphi\|^2,$$

where σ is the smallest positive eigenvalue of $\mathcal{M}_S$.

This inequality along with (5.6) imply that for any φ whose restriction to U is zero we have

$$\frac{\|\mathcal{L}^{1/2}\varphi\|^2}{\|\varphi\|^2} \geq \sigma, \quad \varphi \in L_2(S),$$

or

$$\|\varphi\| \leq \frac{1}{\sqrt{\sigma}}\|\mathcal{L}^{1/2}\varphi\|, \quad \varphi \in L_2(S), \tag{5.7}$$

and Lemma 5.1 implies the inequality

$$\|\varphi\| \leq \frac{1}{\sigma}\|\mathcal{L}\varphi\|, \quad \varphi \in L_2(S).$$

In other words, S is a Λ-set with $\Lambda = \frac{1}{\sigma}$. Therefore, by Theorem 5.1, $U = V(G)\backslash S$ is a uniqueness set for all signals $f \in PW_\omega(G)$ with $\omega < \sigma$. Theorem is proved. ■

Remark 5.2 *In the next Theorem we obtain a similar estimate for an "optimal" set of Paley-Wiener class. It should be noted that this estimate is different from the previous one. It seems, that it is not easy to compare them in general situation.*

Theorem 5.3 *Let $U \subset V(G)$ and $S = V(G)\backslash U$. Let $\mathcal{N}$ be a matrix which is obtained from the matrix of $\mathcal{L}^2$ by replacing by zero columns and rows corresponding to the set U.*

Then U is a uniqueness set for all signals $f \in PW_{\sqrt{\sigma}}(G)$ where σ is the smallest positive eigenvalue of $\mathcal{N}$.

Proof. The proof is similar to the previous one but now for $\varphi \in L_2(S)$ we have the following:

$$\frac{\|\mathcal{L}\varphi\|^2}{\|\varphi\|^2} = \frac{\langle \mathcal{L}\varphi, \mathcal{L}\varphi\rangle}{\|\varphi\|^2} = \frac{\langle \mathcal{L}^2\varphi, \varphi\rangle}{\|\varphi\|^2} = \frac{\langle \mathcal{N}\varphi, \varphi\rangle}{\|\varphi\|^2}. \tag{5.8}$$

Again, the Parseval equality implies the following inequality:

$$\langle \mathcal{N}f, f\rangle \geq \sigma\|f\|^2,$$

where σ is the smallest positive eigenvalue of $\mathcal{N}$.

This inequality along with (5.8) imply that for any φ whose restriction to U is zero we have

$$\frac{\|\mathcal{L}\varphi\|^2}{\|\varphi\|^2} \geq \sigma, \quad \varphi \in L_2(S),$$

or

$$\|\varphi\| \leq \frac{1}{\sqrt{\sigma}}\|\mathcal{L}\varphi\|, \quad \varphi \in L_2(S).$$

In other words, S is a Λ-set with $\Lambda = \frac{1}{\sqrt{\sigma}}$. Therefore, by Theorem 5.1, $U = V(G)\backslash S$ is a uniqueness set for all signals $f \in PW_\omega(G)$ with $\omega = \sqrt{\sigma}$. The theorem is proved. ■

Algorithm to compute the cut-off frequency ω

Given a weighted finite graph $G = (V, E, w)$ and a subset $U \subset V(G)$, let $S = V(G)\backslash S$.

- Compute the Laplacian operator $\mathcal{L}$.
- Compute the submatrix $\mathcal{M}$ of $\mathcal{L}$ containing only the rows and columns corresponding to S.
- Compute σ as the smallest positive eigenvalue of $\mathcal{M}$.

σ is the required cut-off frequency and U is a uniqueness set for all signals $f \in PW_\omega(G)$ with $\omega < \sigma$.

The above theorem proves not only the existence of such a Λ-set but also effectively computes an *"optimal cut-off frequency"* such that the construction is exact if the original graph signal is bandlimited to this frequency. In section 5.3, we will develop another method of reconstruction of bandlimited signals given a graph and the set of vertices for which the signal values are known. We will compute an optimal cut-off frequency and reconstruct signals bandlimited to this frequency using this method.

Note that a more general version of Theorem 5.2 was proved in chapter 4, Theorem 4.7. The theorem shows that for any graph (finite or infinite), any finite subset of the vertex set is a Λ-set. However, the proof requires construction of a more complicated graph from which one needs to compute the corresponding Λ.

Theorem 5.4 *If $PW_\omega(G)$ is finite dimensional for an $\omega > 0$, then a set of vertices $U \subset V(G)$ is a uniqueness set for the space $PW_\omega(G)$ if and only if there exists a constant C_ω such that for any $f \in PW_\omega(G)$ the following discrete version of the Plancherel-Polya inequalities holds true:*

$$\left(\sum_{u\in U} |f(u)|^2\right)^{1/2} \leq \|f\|_{L_2(G)} \leq C_\omega \left(\sum_{u\in U} |f(u)|^2\right)^{1/2} \tag{5.9}$$

for all $f \in PW_\omega(G)$.

Proof: The closed linear subspace $PW_\omega(G)$ is a Hilbert space with respect to the norm of $L_2(G)$. At the same time since $U \subset V(G)$ is a uniqueness set for $PW_\omega(G)$ the functional

$$|||f||| = \left(\sum_{u\in U} |f(u)|^2\right)^{1/2}$$

defines another norm on $PW_\omega(G)$. Indeed, the only property which should be verified is that the condition $|||f||| = 0, f \in PW_\omega(G)$, implies that f is identically zero on the entire graph but it is guaranteed by the fact that U is a uniqueness set for $PW_\omega(G)$.

Since for any $f \in PW_\omega(G)$ the norm $|||f|||$ is not greater than the original norm $\|f\|_{L_2(G)}$, the closed graph Theorem implies our Theorem. ■

5.2 Frames in the spaces $PW_\omega(G)$

Consider a finite graph G with $|V(G)| = N$. In the space $L_2(G)$, consider the kernel

$$K(v,w) = \sum_{0\leq j\leq N-1} e_{\lambda_j}(v)e_{\lambda_j}(w), \qquad (v,w) \in V(G) \times V(G). \tag{5.10}$$

If δ_v is the delta function supported at v, then

$$K(v,w) = \sum_{0 \le j \le N-1} e_{\lambda_j}(v) e_{\lambda_j}(w) = \delta_v(w).$$

In the subspace $PW_\omega(G)$ of $L_2(G)$, consider the orthogonal projection of K onto subspace $PW_\omega(G)$,

$$K_\omega(v,w) = \sum_{0 \le \lambda_i \le \omega} e_{\lambda_i}(v) e_{\lambda_i}(w), \qquad (v,w) \in V(G) \times V(G). \tag{5.11}$$

For a fixed v, this function in w is bandlimited to $[0,\omega]$. More generally, if $h(x)$ on $\mathbb{R}_+$ is supported on $[0,\omega]$, $\omega > 0$, then

$$K^h_\omega(v,w) = \sum_{0 \le \lambda_i \le \omega} h(\lambda_i) e_{\lambda_i}(v) e_{\lambda_i}(w), \qquad (v,w) \in V(G) \times V(G) \tag{5.12}$$

is bandlimited to $[0,\omega]$. Note that, kernel (5.11) corresponds to h which is characteristic function of $[0,\omega]$. Now, given the subspace $PW_\omega(G)$ and a finite sampling set $U \subset V(G)$, K_ω will be used for reconstruction of functions in subspaces $PW_\omega(G)$ and will be called *sampling functions*.

Given $0 \le \omega \le \lambda_{N-1}$, let $U \subset V(G)$ is a sampling set for $PW_\omega(G)$. The following formula will be used for reconstruction of a function $f \in PW_\omega(G)$ through its values on U:

$$\widetilde{f}(v) = \rho \sum_{w \in U} K_\omega(v,w) f(w), \quad v \in V(G), \tag{5.13}$$

where $\rho = \rho(U) = \frac{|V(G)|}{|U|}$ is a scale factor which depends only on the ratio of the cardinalities of V and U. Moreover, the sampling formula (5.13) is exact for bandlimited functions, but it can also be used as an approximation formula for more general functions.

Theorem 5.5 *If G is a finite graph and $U = \{s\} \subset V(G)$ is a sampling set for $PW_\omega(G)$, then the set of functions $K_\omega(s,\cdot) \in PW_\omega(G)$, $s \in U$ is a frame*

in $PW_\omega(G)$. In other words , there exists a constant c_ω such that for every $f \in PW_\omega(G)$,

$$c_\omega \|f\|^2 \leq \sum_{s\in U} |\langle f, K_\omega(s,\cdot)\rangle|^2 \leq \|f\|^2, \tag{5.14}$$

where

$$\langle f, K_\omega(s,\cdot)\rangle = \sum_{v\in V(G)} f(v)K_\omega(s,v), \quad \|f\|^2 = \sum_{v\in V(G)} |f(v)|^2.$$

Proof: For $f \in PW_\omega(G)$, one has

$$\langle f, K_\omega(s,\cdot)\rangle = f(s),$$

and then

$$\sum_{s\in U} |\langle f, K_\omega(s,\cdot)\rangle|^2 = \sum_{s\in U} |f(s)|^2 \leq \sum_{v\in V(G)} |f(v)|^2 = \|f\|^2.$$

The opposite inequality follows from Theorem 5.4 and this completes the proof. ■

At this point we have several ways for reconstruction of ω-bandlimited component of a signal:

1. Since U is the uniqueness (sampling) set then one can use variational splines as a tool for reconstruction of ω-bandlimited component of a signal (see Theorem 5.1 above).

2. Since functions $K_\omega(u,\cdot),\ u \in U$, form frame in the space $PW_\omega(G)$ one can construct canonical dual frame and use it for reconstruction of the ω-bandlimited component of a signal (see Theorem 5.5 above).

3. Since functions $K_\omega(u,\cdot),\ u \in U$, form frame in the space $PW_\omega(G)$ one can use the frame algorithm (see section 2.6). The very first step of this algorithm takes the form

   $$f_1(\cdot) = \lambda \sum_{u\in U} \langle f(\cdot), K_\omega(u,\cdot)\rangle\, K_\omega(u,\cdot) = \lambda \sum_{u\in U} \left(\sum_{v\in U} f(v)K_\omega(u,v) \right) K_\omega(u,\cdot).$$

   It was noticed that for all our examples even this first step gives a satisfactory reconstruction.

5.3 Multiresolution

Given a proper subset of vertices $U \subset V(G)$, its vertex boundary bU is the set of all vertices in $V(G)$ which are not in U but adjacent to a vertex in U:

$$\partial U = \{v \in V(G) \backslash U : \exists \{u, v\} \in E(G), u \in U\}.$$

If a graph $G = (V(G), E(G))$ is connected and U is a proper subset of $V(G)$, then the vertex boundary ∂U is not empty.

For any U which is a subset of vertices of G we introduce the following sequence of subsets of the vertex set $V(G)$.

$$cl^0(U) = U,\ cl(U) = U \cup \partial U,\ cl^m(U) = cl\left(cl^{m-1}(U)\right) = V(G),\ m \in \mathbb{N}. \quad (5.15)$$

Note that

$$cl^0(U) \subseteq cl(U) \subseteq \cdots \subseteq cl^m(U) = V(G),\ m \in \mathbb{N},$$

where m is the smallest positive integer such that $cl^m(U) = V(G)$.

By using Theorem 5.2, we compute the cut-off frequency $\omega_\jmath$ for each subset $cl^\jmath(U),\ \jmath = 0, 1, 2, \ldots, m$. Now, for each function $f \in L_2(G)$ and the set $cl^\jmath(U)$, we define

$$\widetilde{f}_\jmath(v) = \rho_\jmath \sum_{s \in cl^\jmath(U)} K_{\omega_\jmath}(v, s) f(s), \quad v \in V(G), \quad (5.16)$$

where $\rho_\jmath = \frac{|V(G)|}{|cl^\jmath(U)|}$.

If f happens to be in the subspace $PW_{\omega_\jmath}(G)$, then (5.13) guarantees that $\widetilde{f}_\jmath$ is the desired approximation of f. On the other hand, if f does not belong to $PW_{\omega_\jmath}(G)$, we continue the above procedure until we get a satisfactory approximation. Since the operator $\mathcal{L}$ is bounded, every function in $L_2(G)$ belongs to a certain Paley-Wiener space $PW_\omega(G)$ for some $\omega \in \sigma(\mathcal{L})$, where $\sigma(\mathcal{L})$ is the spectrum of $\mathcal{L}$. Hence this process stops after a finite number of steps.

We also notice a progress of reconstruction (get a better approximation) as $\jmath$ increases.

Conclusion: We have developed new sampling techniques and a new method of signal approximation using frames. We have shown that for every subset U of the vertex set V(G), there exists some ω such that U is a sampling set for $PW_\omega(G)$. Thus, the problem of signal interpolation can be posed as the problem of first defining the set of nodes with known sample values as a uniqueness set U, then identifying the maximum ω such that U is a uniqueness set for $PW_\omega(G)$, and then reconstructing the signal values on the complement set $V(G)\backslash U$. A multiresolution analysis on weighted graphs was also discussed. In this case a sequence of sampling sets were constructed and corresponding cut-off frequencies ω's were computed and $PW_\omega(G)$ spaces were constructed. Given a non-bandlimited function, it was shown that its approximation is improving when the set of samples is increasing.

Algorithm to reconstruct any function using multiresolution

Given a weighted finite graph $G = (V, E, w)$, a subset $U \subset V(G)$, and any function $f \in L_2(G)$

- compute the Laplacian operator $\mathcal{L}$.
- compute $\mathcal{L}^2$.
- construct a sequence of subsets $cl^k(U)$ as in (5.15)
- compute the submatrix $\mathcal{M}_k$ of $\mathcal{L}^2$ containing only the rows and columns corresponding to the complement of $cl^k(U)$.
- compute σ_k as the smallest positive eigenvalue of $\mathcal{M}_k$. Then $\sqrt{\sigma}_k$ is the required cut-off frequency and $cl^k(U)$ is a uniqueness set $PW_{\sqrt{\sigma}_k}(G)$.
- construct $\widetilde{f}_k$ from $cl^k(U)$ by using (5.16) and repeat the process (computing $\widetilde{f}_k$) until a desired level of accuracy is achieved.

5.4 Some Examples and Plots

(**1**). In the following figure (fig 5.1) we plotted the 5th Laplacian eigenfunction (blue) and its approximation (red) on cycle graph. We considered a cycle graph with 200 nodes and we took only 5% sampling set which are uniformly distributed throughout the graph. We computed the cut-off frequency and found $\omega = 0.0123$. Since $\lambda_5 = 0.0020 < \omega$, the 5th Laplacian eigenfunction (the eigenvector corresponding to the 5th Laplacian eigenvalue) is indeed in the space $PW_\omega(G)$. We used formula (5.13) to reconstruct f and we can see that the reconstruction is perfect.

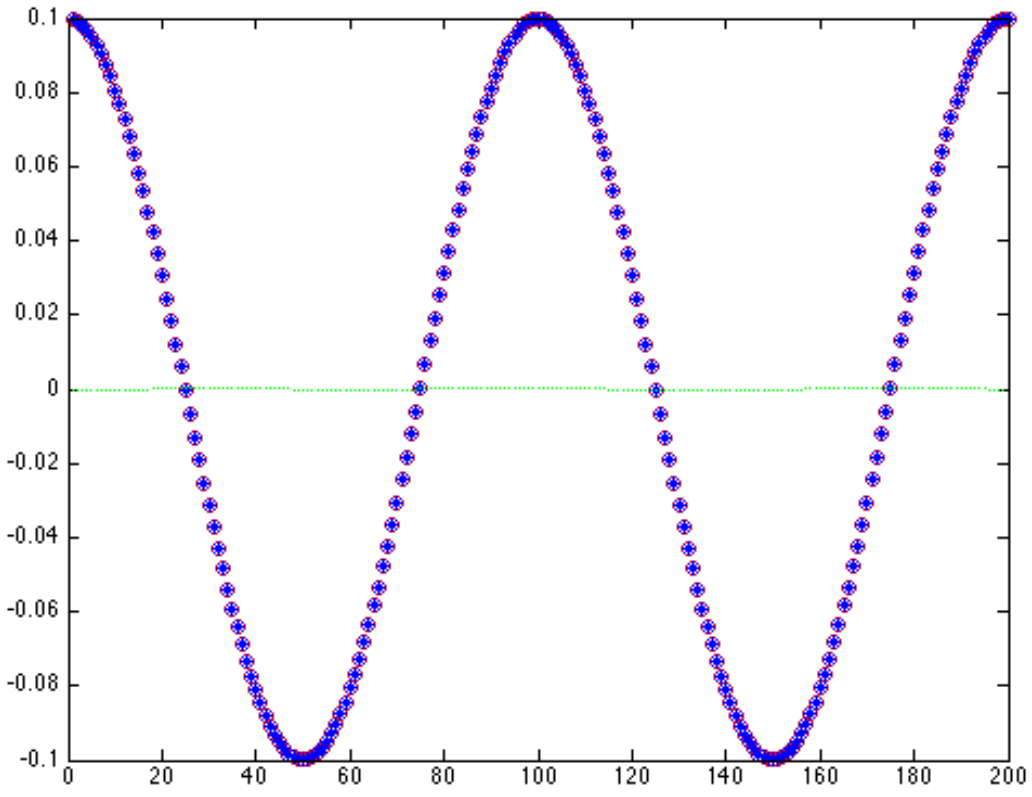


Figure 5.1: 2D approximation (red) of the 5th Laplacian eigenfunction (blue) on a cycle graph of 200 nodes

(**2**). A linear combination of the first 9 eigenfunctions of the Laplacian matrix (blue) and its approximation (red) are plotted below (fig 5.2). We considered the same graph and the same sampling set (only 5%) as in the first case. The coefficients in the linear combination are chosen randomly. This new function is also in the subspace $PW_\omega(G)$, $\omega = 0.0123$ and the reconstruction is perfect as expected.

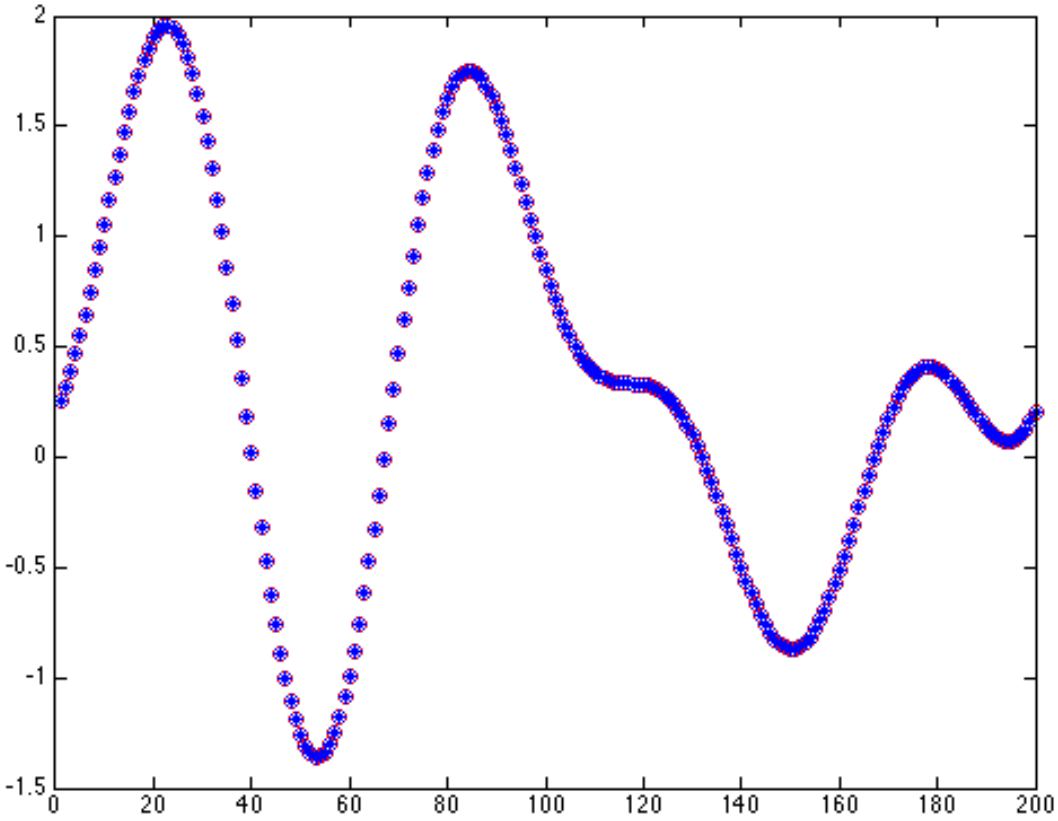


Figure 5.2: 2D approximation (red) of a linear combination of the first 9 eigenfunctions (blue) on a cycle graph of 200 nodes

(**3**). This time we considered a function which is not ω-bandlimited but nearly ω-bandlimited for the same ω. In this case the reconstruction is not perfect but still very good.

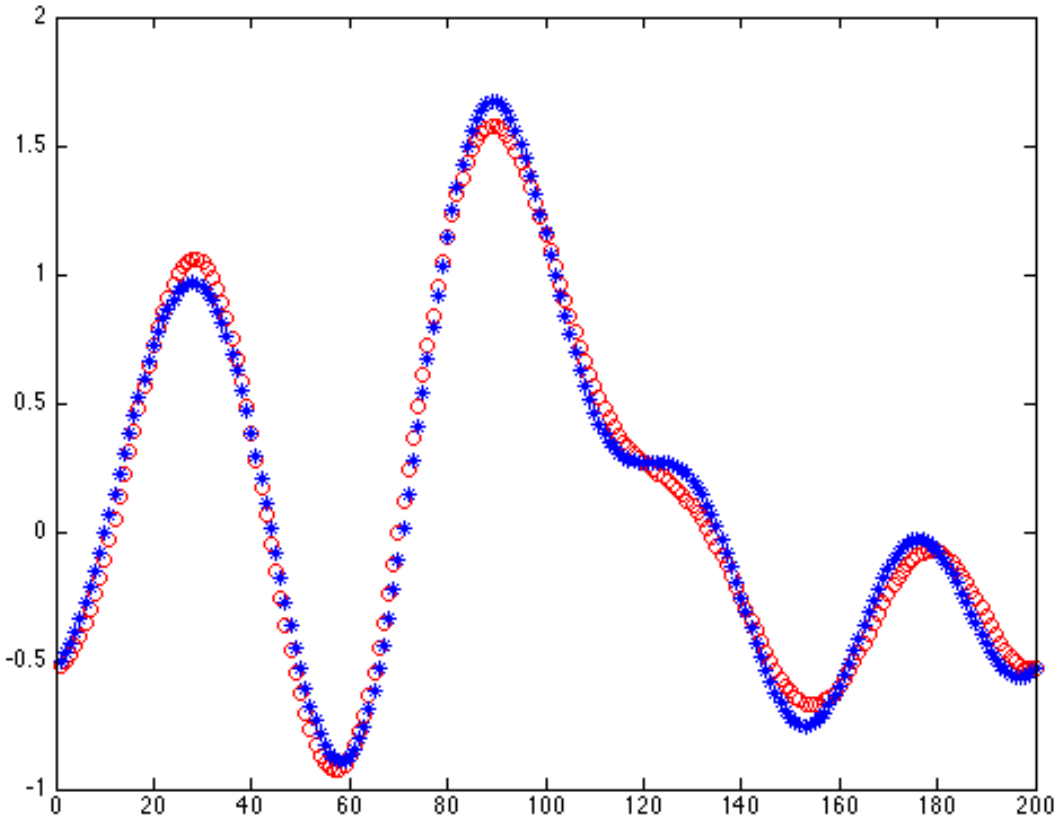


Figure 5.3: 2D approximation (red) of a nearly bandlimited signal (blue) on a cycle graph of 200 nodes

(**4**). In figure 5.4, the 10th Laplacian eigenfunction of a 20 by 20 grid is plotted in 3D. We considered a 20 by 20 grid with 400 nodes and we took only 10% sampling set which is uniformly distributed on the graph. We computed the cut-off frequency and found $\omega = 0.5066$. Since $\lambda_{10} = 0.0585 < \omega$, the 10th Laplacian eigenfunction is in the space $PW_\omega(G)$ and hence ω-bandlimited. So the reconstruction is perfect as shown on the graph.

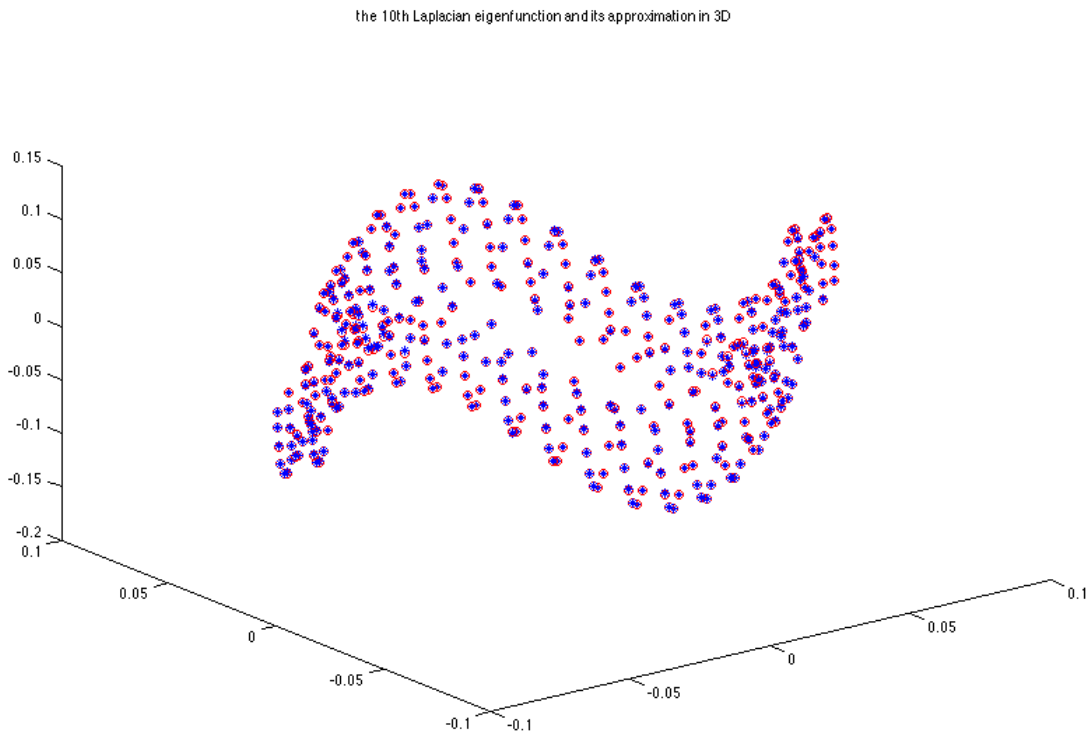


Figure 5.4: 3D approximation (red) of the 10th Laplacian eigenfunction (blue) on a 20 by 20 grid

(**5**). Figure 5.5 is the graph of the same function in 2D. The green curve in this figure shows the difference between the original function and its approximation.

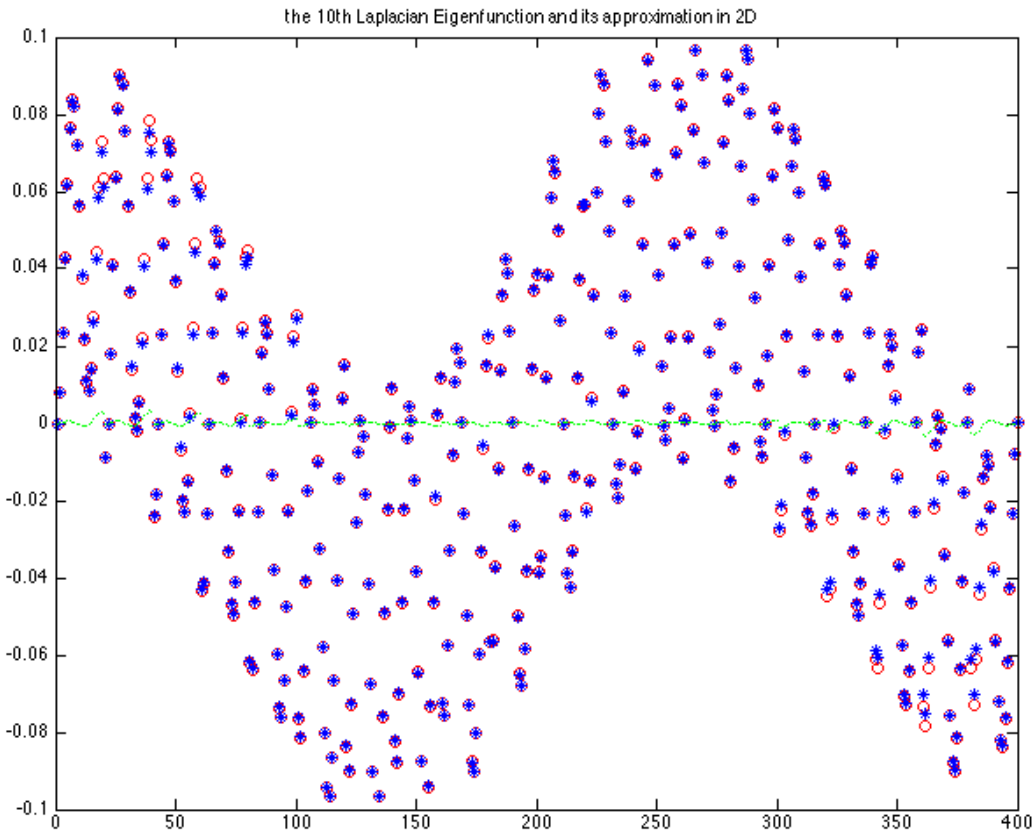


Figure 5.5: 2D approximation (red) of the 10th Laplacian eigenfunction (blue) on a 20 X 20 grid

(**6**). Figure 5.6 and 5.7 show the graphs of kernels K_ω in 3D, for $\omega = \lambda_{10}$ and $\omega = \lambda_{25}$, the 10th and 25th Laplacian eigenvalues on a cycle graph of 200 nodes and figure 5.8 is the graph of kernel K on the same graph.

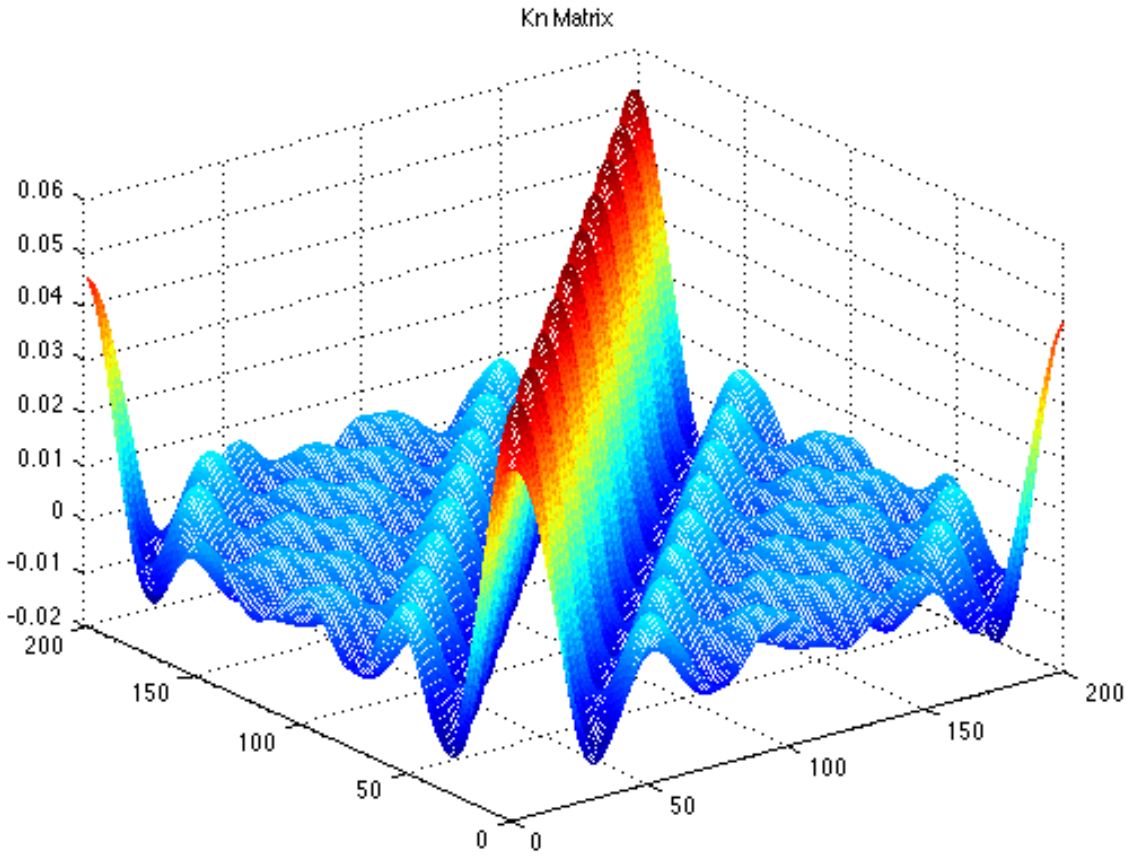


Figure 5.6: Graph of the kernel K_{10} on a 20 by 20 grid

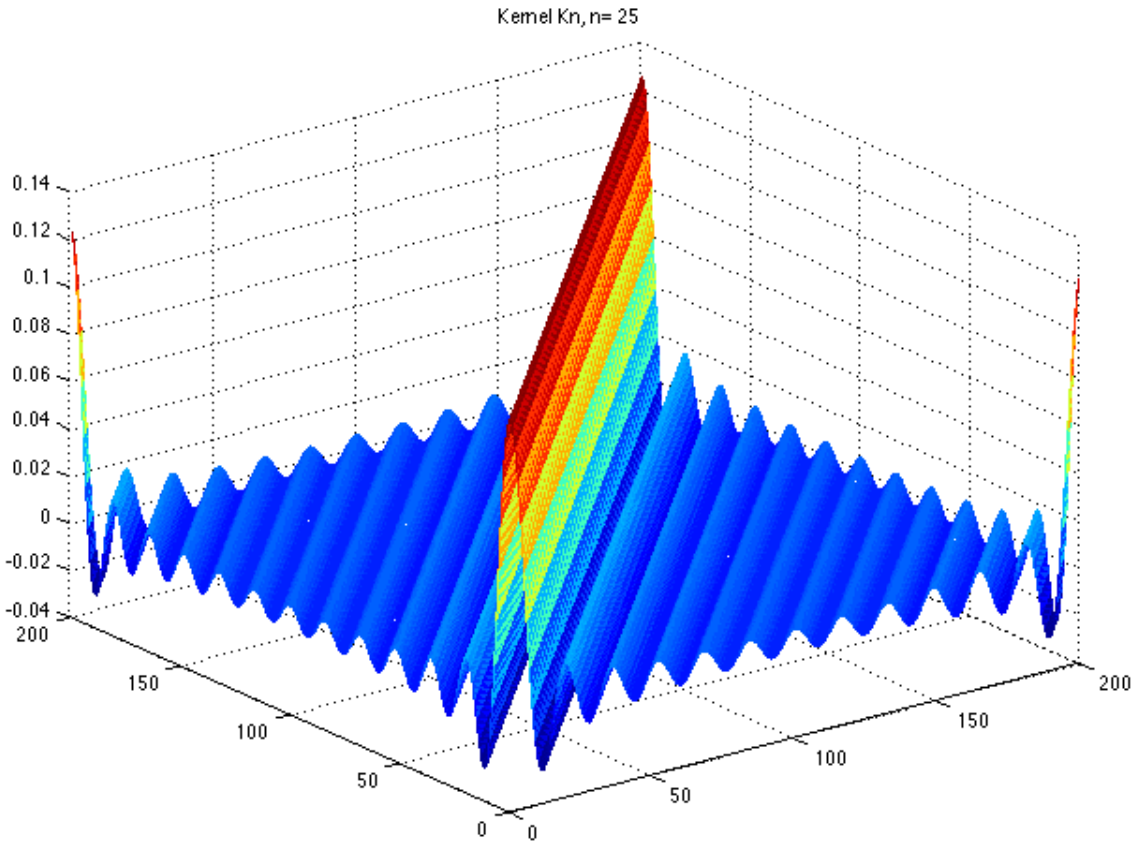


Figure 5.7: Graph of the kernel K_{25} on a 20 by 20 grid

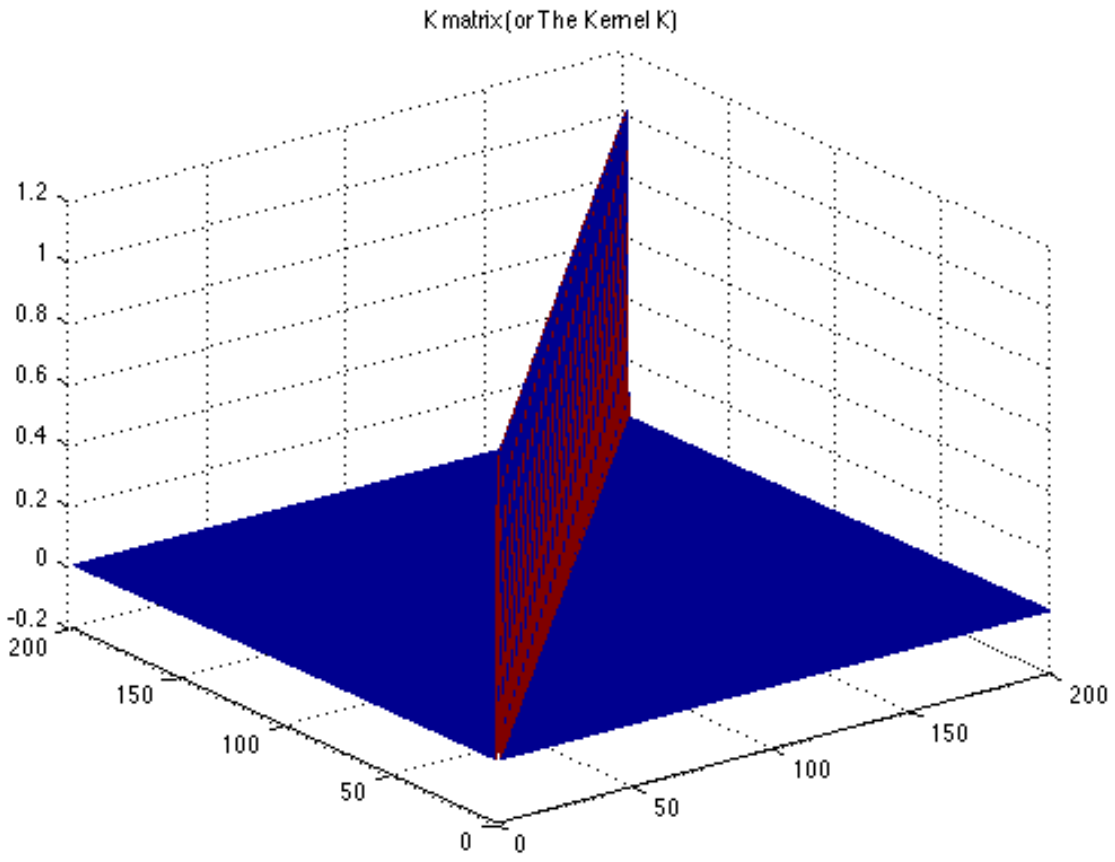


Figure 5.8: Graph of the kernel K on a 20 by 20 grid

CHAPTER 6

SIGNAL APPROXIMATION USING AVERAGE SPLINES

In mathematical analysis, interpolation is a method of constructing new data points from a discrete set of known data points. One often has a number of data points, as obtained by sampling or experiments, and tries to construct a function which closely fits those data points. In other words, interpolating data consists in constructing new values for missing data in coherence with a set of known data. For example, many tasks in image processing and computer vision can be formulated as an interpolation problems [74]; image and video colorization [80], semi-supervised segmentation [81] are examples of these problems.

A different problem which is closely related to interpolation is the approximation of a complicated function from some sample values. Suppose we know the function but it is too complex to evaluate efficiently. Then we could pick a few known data points (sample values) from the complicated function and try to interpolate those data points to construct a simpler function. It is very common to assume that a function is bandlimited, meaning that its natural spectral expansion only lives on a bottom interval of the spectrum. Then an explicit formula, the *sampling formula*, gives the function in terms of its

sample values and certain sampling functions. Generally we may write

$$f(x) = \sum_{y \in S} f(y)\Psi_y(x) \quad \text{for } f \in \mathfrak{B}. \tag{6.1}$$

Here S is the sampling set, $\mathfrak{B}$ is the space of bandlimited functions, and Ψ_y is a sampling frame in the space of bandlimited functions.

In the classical Shannon sampling theorem, the underlying space is the real line, the bandlimited functions $\mathfrak{B}_\omega$ are the L^2 functions whose Fourier transform is supported in an interval $|\lambda| \leq \omega$,

$$f(x) = \int_{-\omega}^{\omega} e^{-2\pi i\lambda} \widehat{f}(\lambda)\, d\lambda : \tag{6.2}$$

S is the lattice $\omega^{-1}\mathbb{Z}$, and $\Psi_y(x)$ is a translate of the sinc function (when $\omega = 1$),

$$\operatorname{sinc} x = \frac{\sin \pi x}{\pi x}.$$

Of course, it is obvious that when using the simple function to calculate new data points we usually do not obtain the same result as when using the original function, but depending on the problem domain and the interpolation method used the gain in simplicity might offset the error.

The theory of splines grew out of the study of simple variational problems. A spline was defined as a function that minimized some notion of energy subject to a set of interpolation constraints. A number of methods for obtaining spline approximations to a given function, or more precisely, to data obtained by sampling a function, have been developed.

In the last few years, the theory of splines was extended to graphs [23]. The theory developed in [23] shows that signals with low frequencies (bandlimited signals) can be nicely approximated by some simple function called *variational splines*. These simple functions interpolate the signal point-wise on a relatively small subset (sampling set) the vertex set of the graph. In this chapter, we will extend the idea to sampling with average values on some subsets of the graph on more general weighted graphs. Variational splines are defined as minimizers of Sobolev norms which are introduced in terms of a combinatorial Laplace

Operator. It is shown that such splines interpolate functions on some subsets of the graph and provide an *optimal* approximations to them. Splines were used as a tool for reconstruction of bandlimited signals from their sampling sets.

6.1 Approximating signals on graphs

In the traditional realm of signal processing, one is interested in approximating a certain function by a simpler one and approximation theory has been well developed. It is well known that the Fourier transform plays an important role in this area. Moreover, the idea that any arbitrary periodic function can be represented as a series of harmonically related sinusoids has a profound impact in mathematical analysis, physics and engineering. In signal processing it has been shown that a smooth signal can be well approximated by a small portion of its Fourier coefficients because of compressibility [44]. It will be shown in this chapter that it is the case for smooth signal supported on graphs.

While a lot has been done on sampling and approximating continuous functions, a considerably less work has been considered on approximating signals supported on graphs. In chapter 5, we have seen a method of approximating signals supported on graphs using the so-called kernels. In the present chapter we will discuss another method of approximating signals using splines and review the work of [23], which defines variational splines on graphs as signals defined on graphs that minimizes some functional defined on the space of functions $L_2(G)$. It is shown that signals from Paley-Wiener spaces can be approximated by variational splines which interpolate the signal on some relatively small subset of the vertices. We will also present results from numerical experiments.

6.1.1 Approximating signals using variational splines

Consider a simple connected weighted finite graph $G = (V, E)$. Let $L_2(G)$ denote the Hilbert space of all real-valued functions $f : V \to \mathbb{R}$. Let L be the Laplacian operator on G and for all positive ω, let $\mathfrak{B}_\omega = \mathfrak{B}_\omega(L)$ be a linear span of the Laplacian eigenfunctions whose eigenvalues are not larger than ω. So $\mathfrak{B}_\omega$ is the space of ω-bandlimited functions.

Variational Problem: Given a subset of vertices $\mathcal{U} \subset V(G)$, a sequence of real numbers $y = \{y_u\}_{u\in\mathcal{U}}$, positive real numbers $k, \epsilon > 0$, we consider the following variational problem.
Find a signal Y from the space $L_2(G)$ which has the following properties:

1. $Y(u) = y_u, \forall u \in \mathcal{U}$,

2. Y minimizes the functional $f \to ||\epsilon I + L)^k f||$.

It is shown in [23] that this variational problem has a unique solution. This unique solution is called the *variational spline* of order k and denoted by $Y^{\mathcal{U},y}_{k,\epsilon}$. The set of all such splines for a fixed $\mathcal{U} \subset V(G)$ and fixed $k, \epsilon > 0$, will be denoted by $\mathfrak{S}(\mathcal{U}, k, \epsilon)$.

Definition 6.1 *We say $L^{\mathcal{U},u}_{k,\epsilon}$ is the Lagrangian spline if it is a solution of the same variational problem with constraints $L^{\mathcal{U},u}_{k,\epsilon}(v) = \delta_{u,v},\ v, u \in \mathcal{U}$, where $\delta_{u,v}$ is the Kronecker delta.*

Theorem 6.1 *[23] For every subset of vertices $\mathcal{U} \subset V(G)$, all $k > 0, \epsilon > 0$, and any sequence of real numbers $y = \{y_u\}_{u\in\mathcal{U}}$,*

1. *the variational problem has a unique solution, $Y^{\mathcal{U},y}_{k,\epsilon}$,*

2. *the solution $Y^{\mathcal{U},y}_{k,\epsilon}$ has a representation*

$$Y^{\mathcal{U},y}_{k,\epsilon} = \sum_{u\in\mathcal{U}} y_u L^{\mathcal{U},u}_{k,\epsilon}. \tag{6.3}$$

Moreover, for any random $\mathcal{U} \subset V(G)$, the set of all Lagrangian splines $L^{\mathcal{U},u}_{k,\epsilon},\ u \in \mathcal{U},\ L^{\mathcal{U},u}_{k,\epsilon}(v) = 0$, if $v \in \mathcal{U}$ and $v \neq u$ is a frame in $\mathfrak{S}(\mathcal{U}, k, \epsilon)$.

Given a function $f \in L_2(G)$, we say that the spline $Y^{\mathcal{U},f}_{k,\epsilon}$ *interpolates* f on $\mathcal{U} \subseteq V(G)$ if $Y^{\mathcal{U},f}_{k,\epsilon}(u) = f(u)$ for all $u \in \mathcal{U}$. It is known that if $f \in L_2(G)$ is bandlimited, then its interpolating spline is always an *optimal approximation*. According to [23], Theorems 1.1 and 3.1, for every subspace $\mathfrak{B}_{m_j}$ one can find a sampling set of points $\mathcal{U}_{m_j} \subset V(G)$ such that every $f \in \mathfrak{B}_{m_j}$ is completely determined by its values on $\mathcal{U}_{m_j}$, where $\mathfrak{B}_{m_j}$ is the subspace of functions bandlimited to $[0, m_j]$. Moreover, every $f \in \mathfrak{B}_{m_j}$ can be reconstructed from its set of values $\{f(s)\}_{s\in\mathcal{U}_{m_j}}$ as a limit of variational splines.

Theorem 6.2 *[23]*

1. *Assume that L is invertible in $L_2(G)$. If S is a Λ-set , then any $f \in \mathfrak{B}_\omega$ with $\omega < \frac{1}{\Lambda}$ can be reconstructed from its values on $\mathcal{U} = V\backslash S$ as the following limit:*

$$f = \lim_{k\to\infty} Y^{\mathcal{U},f}_k, \quad k = 2^\ell, \quad \ell \in \mathbb{N}, \tag{6.4}$$

where $Y^{\mathcal{U},f}_k$ is a spline interpolating f in the set $\mathcal{U} = V\backslash S$ and the error estimate is

$$\|f - Y^{\mathcal{U},f}_k\| \leq 2\gamma^k \|f\|, \quad \gamma = \Lambda\omega < 1. \tag{6.5}$$

2. *If the operator L is not invertible in $L_2(G)$, then for any Λ-set S and any $0 < \epsilon < \frac{1}{\Lambda}$, every function $f \in \mathfrak{B}_\omega$ where $0 < \omega < \frac{1}{\Lambda} - \epsilon$ can be reconstructed from its values on $\mathcal{U} = V\backslash S$ as the following limit:*

$$f = \lim_{k\to\infty} Y^{\mathcal{U},f}_{k,\epsilon}, \qquad k = 2^\ell, \quad \ell \in \mathbb{N}, \tag{6.6}$$

where $Y^{\mathcal{U},f}_{k,\epsilon}$ is a spline interpolating f in the set $\mathcal{U} = V\backslash S$ and the error estimate is

$$\|f - Y^{\mathcal{U},f}_{k,\epsilon}\| \leq 2\gamma^k \|f\| \quad \gamma = \Lambda(\omega + \epsilon) < 1. \tag{6.7}$$

Algorithm to construct approximating splines

1. Pick a random subset of vertices $S \subset V(G)$
2. Construct Lagrangian splines with nodes in S using eq.(6.3)
3. Use the algorithm in chapter 5 to find an optimal frequentcy ω such that S is a uniqueness set for $\mathfrak{B}_\omega$.
4. Use spline interpolation for reconstruction of functions in $\mathfrak{B}_\omega$ from their values on S.

Example 1. Consider a cycle graph $C_{100} = \{1, \ldots, 100\}$. In figure 6.1 we plot the graph of the 6th Laplacian eigenfunction(red) and its approximation using splines (blue) on a cycle graph of 100 nodes. We took only 10% of the vertices as our sampling set and $k = 15$. The green curve is the difference (in absolute value) of the original function and its approximation.

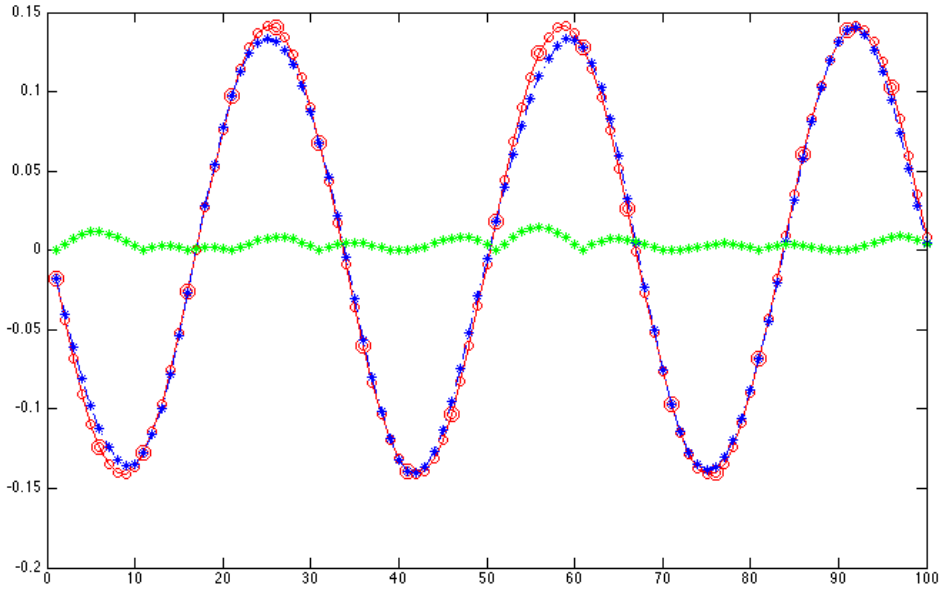


Figure 6.1: Approximating the 6th eigenfunctions (red) by its interpolating variational splines (blue)

Example 2. In this example we consider the 6th Laplacian eigenfunction on a cycle graph of 200 nodes. We took only 5% of the nodes as our sampling set (uniformly distributed). Then we reconstruct the eigenfunction using the method of chapter 5 and compare it with the previous example. In this case the reconstruction is exact as opposed to the previous example even with a smaller sampling set.

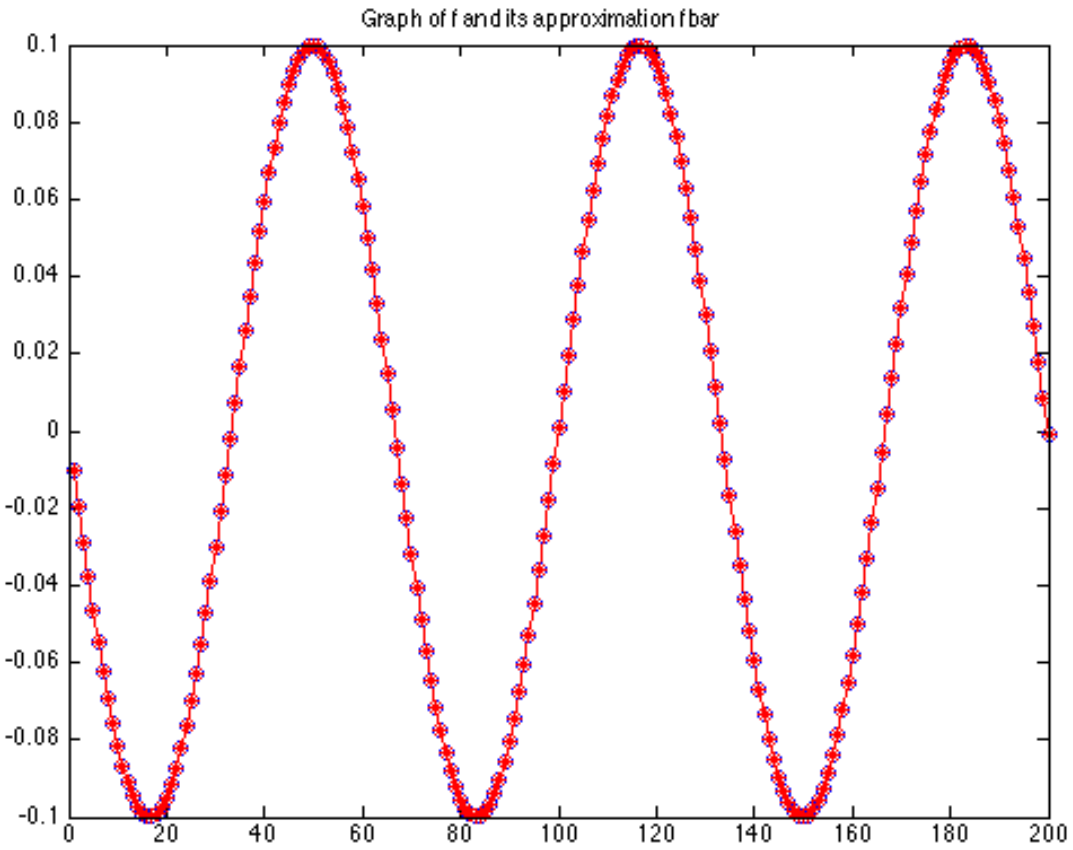


Figure 6.2: Approximating the 6th eigenfunctions (red) by kernels (blue)

Example 3. This time we consider the Delaunay graph of 400 randomly chosen points in the unit square. We then chose 30% of uniformly distributed points and approximate the 10th Laplacian eigenfunctions.

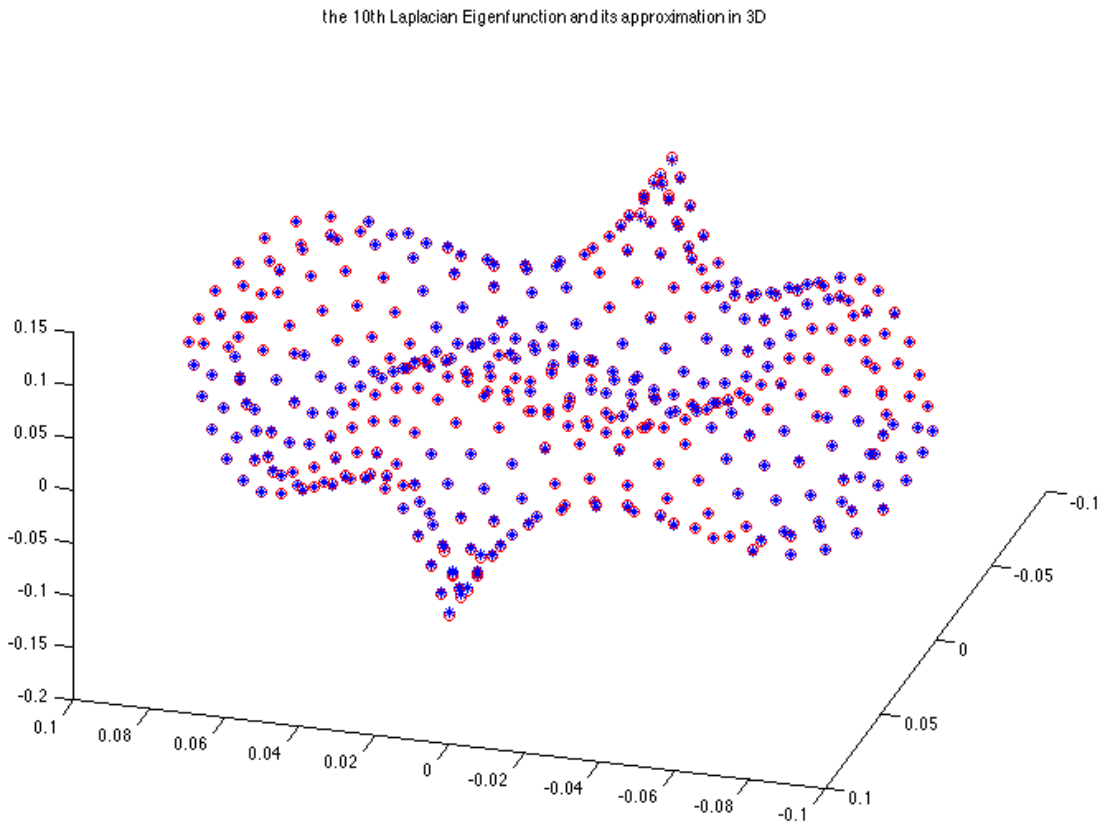


Figure 6.3: the 10th Laplacian eigenfunction (red) of the Delaunay graph of 400 nodes and is approximation (blue), 30% sampling set.

6.2 Sampling on graphs using average splines

Sampling theory is one of the most powerful results in signal analysis. The objective of sampling is to reconstruct a signal from its samples. For example, if f is bandlimited to $[-\omega, \omega]$, then f is uniquely determined and can be reconstructed by its sample at $x_k = \frac{k\pi}{\omega}$, which is the classical Shannon sampling theorem. In a similar way, we have seen in chapter 5 and section 7.1 that if a graph signal f is bandlimited to $[0, \omega]$, then f is uniquely determined and can be reconstructed from its values on the uniqueness set.

Although the assumption that a signal is bandlimited is eminently useful, it is not always realistic since a bandlimited signal is of finite duration. Moreover, in practice, measured values may not be values of a signal f precisely at times x_k, but only local averages of f near x_k. Specifically, measured sampled values are

$$\langle f, u_k \rangle = \int f(x) u_k(x)\, dx,$$

for some collection of average functions $u_k(x), k \in \mathbb{Z}$, which satisfy the following properties:

$$\operatorname{supp} u_k \subset \left[x_k - \frac{\delta}{2}, x_k + \frac{\delta}{2} \right], \qquad u_k(x) \geq 0, \text{and} \quad \int u_k(x)\, dx = 1.$$

It is clear that from local averages one should obtain at least a good approximation of the original signal if δ is small enough. Wiley, Butzer and Lei studied the approximation error when local averages are used as sampled values [101, 102]. Furthermore, Feichtinger and Grochenig [88, 103] proved that a signal is uniquely determined by its local averages under certain conditions.

In a precise analogy, in some cases sampling with average values on subgraphs is more natural than point-wise sampling. In this chapter, we study the reconstruction of bandlimited graph signals from their average values in some subsets of the graph.

Definition 6.2 *Given a weighted graph $G = (V, E, w)$ and a function $f \in L_2(G)$, we define its* ***average value*** *over a subset of vertices $S \subset V$ as the*

weighted average:

$$\mathcal{F}_S(f) := \frac{\sum_{v\in S} d(v)f(v)}{vol(S)}, \tag{6.8}$$

where

$$d(v) = \sum_{u\sim v} w(u,v) \text{ and } vol(S) = \sum_{v\in S} d(v).$$

In point sampling and interpolation the approximating function (interpolant) and the underlying function must agree at a finite number of points. Of course, in most cases additional constraints may be imposed on the interpolant depending on the problem (for example, smoothness). Furthermore, a choice must be made about which family of functions the interpolant is a member of. For instance, if the underlying function is continuous on some interval of the set of real numbers, we may require our interpolant to be a polynomial (Polynomial Interpolation), or a trigonometric function (Fourier Approximation). On the other hand, in the discrete (graph) setting, we require the interpolant to be a member of the so called "Sobolev space", which can be defined as the domain of the powers of the Laplacian operator [23].

In the present chapter we introduce a new sampling and approximation technique for signals defined on graphs using their average values on disjoint subsets of the vertex set. This time we are generally assuming that the values of the function are not required to be known at each vertex of the graph, but instead its average values are known or at least can be estimated from some observation on some disjoint subsets of the vertex set. As in the case of point sampling, the interpolant is shown to be a linear combination of some set of basis functions for the family of functions chosen for the interpolation (spline spaces).

Variational Problem: Given a disjoint subsets $\{V_j\}_{j=1}^m$ of $V(G)$, a sequence of real numbers $(r_1, r_2, \ldots, r_m)$, and positive real numbers $t, \epsilon > 0$, find a function $s : V \to \mathbb{R}$ which attains the prescribed averages,

$$\mathcal{F}_{V_j}(s) = r_j, \quad j = 1, \ldots, m \tag{6.9}$$

and minimizes the functional

$$f \to ||(\epsilon I + L)^t f||, \tag{6.10}$$

that is

$$||(\epsilon I + L)^t s|| = \min_f ||(\epsilon I + L)^t f||,$$

where min is taken over all functions $f \in L_2(G)$ subject to the average constraints of (6.9) for a prescribed set $\{r_j\}_j$. More generally, given a function $f \in L_2(G)$, we will seek an interpolant $s_t(f) \in L_2(G)$ such that

$$\mathcal{F}_{V_j}(s_t(f)) = \mathcal{F}_{V_j}(f), \text{ for all } 1 \leq j \leq m, \tag{6.11}$$

and $s_t(f)$ minimizes the norm $||(\epsilon I + L)^t g||$, where $g \in L_2(G)$ satisfying (6.11). In this case, we say $s_t(f)$ is an average spline interpolating f over $V_1, \ldots, V_m$. A generalization of [23], and [40], implies that the above variational problem has a unique solution. We call this unique solution an *average interpolating spline of order* t and denote it by $s_t(f)$. Furthermore, the solution of this variational problem can be written in terms of the fundamental solutions of the operator $(\epsilon I + L)^{2t}$:

$$(\epsilon I + L)^{2t} s_t(f) = \sum_{j=1}^{m} \alpha_j \mathcal{F}_{V_j}(f), \tag{6.12}$$

where $\{\alpha_j\}_j = \{\alpha_j(s_t(f))\}_j$ is a sequence of constants which depend on $s_t(f)$.

On $L_2(G)$ we define a scalar product by

$$\langle f, g \rangle_{H_t(G)} = \langle (\epsilon I + L)^{t/2} f, (I + L)^{t/2} g \rangle_{L_2(G)} \quad t > 0, \tag{6.13}$$

and denote the corresponding norm by $||f||_{H_t(G)}$. We then define $H_t(G)$ to be the space of all functions from $L_2(G)$ with the norm $||f||_{H_t(G)} = \langle f, f \rangle_{H_t(G)}$. We note that for any $f \in H_t(G)$,

$$|f||_{H_t(G)} = \left(\sum_{j=0}^{N-1} (\epsilon + \lambda_j)^t |c_j(f)|^2 \right)^{1/2}, \tag{6.14}$$

where $c_j(f)$ are the Fourier coefficients of f defined as follows:

$$c_j(f) = \hat{f}(\lambda_j) = \langle f, \phi_j \rangle = \sum_{v \in V} f(v)\phi_j^*(v), \tag{6.15}$$

and ϕ_j is the orthonormal eigenfunction corresponding to the j^{th} Laplacian eigenvalue λ_j and ϕ_j^* its conjugate transpose.

Since the Laplace operator L is bounded, all the spaces $H_t(G)$ coincide as a set. The average interpolating spline is then the function that minimizes the functional

$$F(f) = ||(\epsilon I + L)^{t/2} f|| \tag{6.16}$$

over the function in the space $H_t(G)$.

Remark 6.1 *Equation (6.12) is satisfied in the sense of distribution, meaning that for any $\psi \in H_t(G)$,*

$$\langle (\epsilon I + L)^t s_t(f), \psi \rangle_{L_2(G)} = \sum_{j=1}^{m} \alpha_j(s_t(f)) \mathcal{F}_{V_j}(\psi). \tag{6.17}$$

From now on, if there is confusion, we will drop the subscript V_j and use the notation $\mathcal{F}_j$ for $\mathcal{F}_{V_j}$ just for the sake of simplicity.

The class of all average splines of order t with interpolating sets $\mathcal{V} = \{V_1, \ldots, V_m\}$ and a fixed $\epsilon > 0$ is denoted by $\mathcal{S}(\mathcal{V}, t, \epsilon)$. Note that if each subset V_j is a point (contains a single vertex), our optimization problem will be reduced to the variational problem defined in [23]. Basically, our formulation of the average interpolation approach is a direct generalization of point-wise interpolation on graphs [23].

Lemma 6.1 *For each $1 \leq j \leq m$, there is a function $\mathfrak{h}_j$ in $\mathcal{S}(\mathcal{V}, t, \epsilon)$ satisfying the property*

$$\mathcal{F}_j(\mathfrak{h}_i) = \delta_{ij}, \tag{6.18}$$

where δ_{ij} is the Kronecker delta. Moreover, these functions form a basis of $\mathcal{S}(\mathcal{V}, t, \epsilon)$.

Proof: Let $\mathfrak{h}_j$ be a solution of the optimization problem corresponding to $\beta = (0, 0, \ldots, 1, 0, \ldots, 0)$, the jth row of an $m \times m$ identity matrix. Clearly we have $\mathcal{F}_i(\mathfrak{h}_j) = \delta_{ij}$, $1 \leq i, j \leq m$.

It is easy to show that these functionals are linearly independent. In fact, if there are scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ for which

$$\sum_{j=1}^{m} \alpha_j \mathcal{F}_j = 0,$$

then for each $1 \leq k \leq m$

$$0 = \sum_{j=1}^{m} \alpha_j \mathcal{F}_j(\mathfrak{h}_k) \ = \sum_{j=1}^{m} \alpha_j \delta_{jk} = \ \alpha_k. \quad \blacksquare$$

Later we will show that for any $t \geq 0$, the space of splines of order t is a finite dimensional linear subspace of $L_2(G)$, with dimension m and basis set $\{\mathfrak{h}_j\}_{j=1}^{m}$.

Now observe that our variational problem can be explained as follows. Given a sequence of real numbers $(r_1, r_2, \ldots, r_m)$, and $t \geq 0$ consider the following minimization problem:
Find a function s from the space $H_t(G)$ satisfying the following properties

1. $\mathcal{F}_j(s) = r_j, \ j = 1, 2, \ldots, m$

2. s minimizes the functional $F(s) = ||(\epsilon I + L)^{t/2} s||$.

Let $\mathbf{r} = (r_1, \ldots, r_m)$ denote any vector in $\mathbb{R}^m$. We denote the set of all functions f from $H_t(G)$ such that $\mathcal{F}_j(f) = r_j$ by $E_t(\mathbf{r})$, where $1 \leq j \leq m$.

That is,

$$E_t(\mathbf{r}) = \{f \in H_t(G) : \mathcal{F}_j(f) = r_j, \ \mathbf{r} = (r_1, \ldots, r_m), \ 1 \leq j \leq m\}. \tag{6.19}$$

In particular, we denote

$$E_t(\mathbf{0}) = \{f \in H_t(G) : \mathcal{F}_j(f) = 0, \ \ 1 \leq j \leq m\} \ \text{ by } E_t^0, \tag{6.20}$$

where $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Note that

$$E_t^0 = \bigcap_{j=1}^{m} Ker \mathcal{F}_j.$$

Theorem 6.3 *(Necessary and Sufficient Conditions) For any sequence of values* $(r_1, r_2, \ldots, r_m) = \mathbf{r}$ *and any* $t \geq 0$*, a function* $f \in E_t(\mathbf{r})$ *is a solution of the variational problem if and only if it is orthogonal to the subspace* E_t^0 *with respect to the inner product in* $H_t(G)$*, that is*

$$\langle f, g \rangle_{H_t(G)} = 0, \forall g \in E_t^0.$$

Proof: Assume $f \in E_t(\mathbf{r})$ and $\langle f, g \rangle_{H_t(G)} = 0, \forall g \in E_t^0$. Then for any constant α and any function $g \in E_t^0$ we have

$$\begin{aligned} ||f + \alpha g||^2_{H_t(G)} &= ||(\epsilon I + L)^{t/2}(f + \alpha g)||^2_{L_2(G)} \\ &= ||(\epsilon I + L)^{t/2} f||^2_{L_2(G)} + |\alpha|^2 ||(\epsilon I + L)^{t/2} g||^2_{L_2(G)} \\ &= ||f||^2_{H_t(G)} + |\alpha|^2 ||g||^2_{H_t(G)}, \end{aligned}$$

which implies f is a minimizer.
Conversely, assume that $f \in E_t(\mathbf{r})$. Let g be any element of E_t^0. Then

$$\begin{aligned} ||f + \alpha g||^2_{H_t(G)} &= ||(\epsilon I + L)^{t/2}(f + \alpha g)||^2_{L_2(G)} \\ &= ||(\epsilon I + L)^{t/2} f||^2_{L_2(G)} + 2\alpha \langle f, g \rangle_{H_t(G)} + |\alpha|^2 ||(\epsilon I + L)^{t/2} g||^2_{L_2(G)} \\ &= ||f||^2_{H_t(G)} + 2\alpha \langle f, g \rangle_{H_t(G)} + |\alpha|^2 ||g||^2_{H_t(G)}. \end{aligned}$$

The function f can be a minimizer only if $\langle f, g \rangle_{H_t(G)} = 0$. This completes the proof. ■

Corollary 6.1 $\mathcal{S}(\mathcal{V}, t, \epsilon) \bigcap E_t^0 = \mathbf{0}$.

Theorem 6.4 *For any sequence of values* $\mathbf{r} = (r_1, r_2, \ldots, r_m)$ *and any* $t \geq 0$*, the variational problem has a unique solution.*

Proof: It is clear from the definition that for any sequence of values $(r_1, r_2, \ldots, r_m) = \mathbf{r}$, the linear manifold $E_t(\mathbf{r})$ is a shift of the closed subspace E_t^0, that is,

$$E_t(\mathbf{r}) = E_t^0 + h,$$

where h is any function from $H_t(G)$ such that $\mathcal{F}_j(h) = r_j,\ j = 1, 2, \ldots, m$.

Let P_h denote the orthogonal projection of $h \in H_t(G)$ onto the closed subspace E_t^0 with respect to the inner product in $H_t(G)$:

$$\langle f, g\rangle_{H_t(G)} = \langle(\epsilon I + L)^{t/2} f, (I+L)^{t/2} g\rangle_{L_2(G)} = \sum_{v \in V(G)} ((I+L)^{t/2} f)(v)((I+L)^{t/2} g)(v).$$

The function $s_t(\mathbf{r}) = h - P_h$ is the unique solution of the variational problem. Observe that $s_t(\mathbf{r}) \in E_t(\mathbf{r})$. So we only need to show that $s_t(\mathbf{r})$ minimizes the functional

$$F(g) = ||(\epsilon I + L)^{t/2} g||$$

on the linear manifold $E_t(\mathbf{r})$. To see this, pick any function $f \in E_t(\mathbf{r})$. Then f can be written in the form $f = s_t(\mathbf{r}) + g$ for some $g \in E_t^0$. Then we have

$$\begin{aligned} ||f||^2_{H_t(G)} &= ||(\epsilon I + L)^{t/2}(s_t(\mathbf{r}) + g)||^2 \\ &= ||s_t(\mathbf{r})||^2_{H_t(G)} + 2\langle s_t(\mathbf{r}), g\rangle_{H_t(G)} + ||g||^2_{H_t(G)}. \end{aligned}$$

But we know that $s_t(\mathbf{r}) = h - P_h$ is orthogonal to E_t^0. Thus for any constant α we have have

$$||s_t(\mathbf{r}) + \alpha g||^2_{H_t(G)} = ||s_t(\mathbf{r})||^2_{H_t(G)} + |\alpha|^2 ||g||^2_{H_t(G)},$$

which implies that the function $s_t(\mathbf{r})$ is the minimizer. ■

Theorem 6.5 *The space of splines $\mathcal{S}(\mathcal{V}, t, \epsilon)$ of any order $t \geq 0$ is a finite dimensional linear subspace of $L_2(G)$, with dim $\mathcal{S}(\mathcal{V}, t) = m$.*

Proof: To show linearity, let s and s' be splines in $\mathcal{S}(\mathcal{V}, t)$ interpolating $\mathbf{r}$ and $\mathbf{r}'$ respectively and let $s^* = \alpha s + \alpha' s'$, where α, α' are scalars. We claim that s^* solves the minimization problem for $\mathbf{r}^* = \alpha \mathbf{r} + \alpha' \mathbf{r}'$.

Indeed, $s^* \in E_t(\mathbf{r}^*)$, and by Theorem 6.3, if $f \in E_t^0$, then we have

$$\begin{aligned} \langle f, s^*\rangle &= \langle f, \alpha s + \alpha' s'\rangle \\ &= \alpha\langle f, s\rangle + \alpha'\langle f, s'\rangle \\ &= 0. \end{aligned}$$

Thus (by Theorem 6.3), s^* is a solution of the variational problem for $\mathbf{r}^*$.

Next, to show $\dim \mathcal{S}(\mathcal{V}, t, \epsilon) = m$, we will show by constructing an explicit basis for $\mathcal{S}(\mathcal{V}, t, \epsilon)$. Let $\mathfrak{h}_j$ be solution of the optimization problem corresponding to $\mathbf{r} = (0, 0, \ldots, 1, 0, \ldots, 0)$, the jth row of an $m \times m$ identity matrix. We have seen that $\mathcal{F}_i(\mathfrak{h}_j) = \delta_{ij}, \quad 1 \leq i, j \leq m$, where δ_{ij} is the Kronecker delta. We claim that $\mathcal{S}(\mathcal{V}, t, \epsilon) = \text{span}\{\mathfrak{h}_j\}_{j=1}^m$. To see this, let $f \in \mathcal{S}(\mathcal{V}, t, \epsilon)$, and let

$$\theta = f - \sum_{j=1}^{m} (\mathcal{F}_j(f))\mathfrak{h}_j.$$

Since both f and each $\mathfrak{h}_j$ are in $\mathcal{S}(\mathcal{V}, t, \epsilon)$, then $\theta \in \mathcal{S}(\mathcal{V}, t, \epsilon)$. Moreover, for each $1 \leq i \leq m$,

$$\begin{aligned}
\mathcal{F}_i(\theta) &= \mathcal{F}_i(f) - \sum_{j=1}^{m} (\mathcal{F}_j(f))\mathcal{F}_i(\mathfrak{h}_j) \\
&= \mathcal{F}_i(f) - \sum_{j=1}^{m} (\mathcal{F}_j(f))\delta_{ij} \\
&= \mathcal{F}_i(f) - \mathcal{F}_i(f) = 0,
\end{aligned}$$

which implies $\theta \in E_t^0$.
Thus, $\theta \in E_t^0 \bigcap \mathcal{S}(\mathcal{V}, t, \epsilon)$, and by Theorem 6.3, we have

$$0 = \langle \theta, \theta \rangle_{H_t(G)} = ||\theta||^2_{H_t(G)},$$

which then implies that $\theta = 0$. Hence,

$$f = \sum_{j=1}^{m} (\mathcal{F}_j(f))\mathfrak{h}_j,$$

which proves that $\{\mathfrak{h}_j\}_{j=1}^m$ spans $\mathcal{S}(\mathcal{V}, t, \epsilon)$.

Now, it remains to show that the set $\{\mathfrak{h}_j\}$ is linearly independent. Suppose that

$$\sum_{j=1}^{m} \alpha_j \mathfrak{h}_j = \mathbf{0},$$

for some scalars α_j. Then for each $1 \leq i \leq m$,

$$0 = \mathcal{F}_i(\mathbf{0}) = \sum_{j=1}^{m} \alpha_j \mathcal{F}_i(\mathfrak{h}_j) = \sum_{j=1}^{n} \alpha_j \delta_{ij} = \alpha_i.$$

Hence, $\{\mathfrak{h}_j\}_{j=1}^m$ is linearly independent and therefore it is a basis for $\mathcal{S}(\mathcal{V}, t, \epsilon)$. ■

The following corollary is a simple consequence of the above theorem and our definition of interpolation.

Corollary 6.2 *Let $f \in L_2(G)$ and $s_t(f) \in \mathcal{S}(\mathcal{V}, t, \epsilon)$ is a spline interpolating f. Then $s_t(f)$ can be explicitly written as*

$$s_t(f) = \sum_{j=1}^{m} (\mathcal{F}_j(f))\mathfrak{h}_j. \tag{6.21}$$

Proof: For any $f \in L_2(G)$, we have $\mathcal{F}_j(s_t(f)) = \mathcal{F}_j(f),\ 1 \leq j \leq m$ and by applying Theorem 6.5, we obtain

$$s_t(f) = \sum_{j=1}^{m} (\mathcal{F}_j(s_t(f)))\mathfrak{h}_j = \sum_{j=1}^{m} (\mathcal{F}_j(f))\mathfrak{h}_j. \quad ■$$

The next theorem shows that a bandlimited signal from $L_2(G)$ can be uniquely determined and reconstructed from its average values as a limit of average splines.

Theorem 6.6 *If $f \in PW_\omega(G)$ and $s_t(f)$ is an average spline interpolating f on some disjoint sets $\{V_j\}_j$, then f is uniquely determined by its average values on these disjoint sets and can be reconstructed as a limit of its average interpolating splines through the formula*

$$f = \lim_{t \to \infty} s_t(f). \tag{6.22}$$

Conclusion: We have introduced a new sampling signal approximation technique using average interpolating splines. The average value of a signal on a subset of the vertex set is defined as a weighted average and an approximation technique is introduced using splines interpolating the signal on disjoint subsets of the vertex set. we have shown that a bandlimited signal from $L_2(G)$ can be uniquely determined and reconstructed from its average values as a limit of average interpolating splines.

CHAPTER 7

QUADRATURES ON COMBINATORIAL GRAPHS

7.1 Introduction

The goal of this chapter is to establish quadratures on combinatorial graphs. Two types of quadratures are developed. Quadratures of the first type are exact on spaces of variational splines on graphs. Since bandlimited functions can be obtained as limits of variational splines we obtain quadratures which are "essentially" exact on spaces of bandlimited functions. Quadratures of the second type are exact on spaces of bandlimited functions. Accuracy of quadratures is given in terms of smoothness which is measured by means of combinatorial Laplace operator. The results have potential applications to problems that arise in data mining.

Quadratures for approximate and exact evaluation of integrals of functions defined on Euclidean spaces or on smooth manifolds is an important and continuously developing subject. In recent years in connection with applications to information theory analysis of functions defined on combinatorial graphs attracted attention of many mathematicians.

In particular, certain quadratures for functions defined on combinatorial graphs were recently considered in the paper [104]. There, given values of a

function f on a small subset U of the set of all vertices V of a graph, the authors estimate wavelet coefficients via specific quadratures.

In the present chapter we develop a set of rules (quadratures) which allow for approximation or exact evaluation of "integrals" $\sum_{v\in V} f(v)$ of functions by using their values on subsets $U \subset V$ of vertices. We make extensive use of the previous work on Shannon sampling of bandlimited functions and variational splines on combinatorial graphs. Our results can find applications to problems that arise in connection with data filtering, data denoising and data dimension reduction.

In section 7.2 by using interpolating splines we develop a set of quadratures. Theorem 7.1 shows that these formulas are *exact* on the set of variational splines. Theorem 7.3 explains that our quadratures are *optimal*.

In section 7.3, using the fact that bandlimited functions are limits of variational splines, we show that quadratures developed in section 7.2 are "essentially" exact on bandlimited functions.

It can be verified, for example, that for a cycle graph of 1000 vertices a set of about 670 "uniformly" distributed vertices is sufficient to have asymptotically exact quadratures for linear combinations of the first 290 eigenfunctions (out of 1000) of the corresponding combinatorial Laplace operator.

It is worth to noting that all results of section 7.2 which provide errors of approximation of integrals of functions on V through their values on a $U \subset V$ reflect

1) geometry of U which is inherited into the quantity $\sqrt{|V|-|U|} = \sqrt{|S|}$ and into the Poincaré constant Λ (see section 7.2 for definitions),

2) smoothness of functions which is measured in terms of combinatorial Laplace operator.

In section 7.4, we develop a different set of quadratures which are *exact* on appropriate sets of bandlimited functions. The results in this section are formulated in the language of frames and only useful if it is possible to calculate dual frames explicitly. Since in general it is not easy to compute a dual frame we finish this section by explaining another approximate quadrature which is

based on the so-called frame algorithm.

7.2 Quadratures which are exact on variational splines

Let us remind ourselves the definition of variational splines.

Variational Problem

Given a subset of vertices $U = \{u\} \subset V$, a sequence of real numbers $\overline{y} = \{y_u\} \in l_2, u \in U$, a natural number k, and a positive $\varepsilon > 0$, we consider the following variational problem.

Find a function Y from the space $L_2(G)$ which has the following properties:

1) $Y(u) = y_u, u \in U$,

2) *Y minimizes functional* $Y \to \left\|(\varepsilon I + \mathcal{L})^k Y\right\|$.

We show that the above variational problem has a unique solution $Y_{k,\varepsilon}^{U,\overline{y}}$.

For the sake of simplicity we will also use the notation $Y_k^{\overline{y}}$ assuming that U and ε are fixed.

It is known that for every set of vertices $U = \{u\}$, every natural k, every $\varepsilon \geq 0$, and for any given sequence $\overline{y} = \{y_u\} \in l_2$, the solution $Y_k^{\overline{y}}$ of the Variational Problem has a representation

$$Y_k^{\overline{y}} = \sum_{u \in U} y_u L_k^u, \tag{7.1}$$

where L_k^u is the so-called Lagrangian spline, i.e., it is a solution of the same Variational Problem with constraints $L_k^u(v) = \delta_{u,v}$, $u \in U$, where $\delta_{u,v}$ is the Kronecker delta.

We introduce the following scalars:

$$\theta_u = \theta_u(U,\, k,\, \varepsilon) = \sum_{v \in V} L_{k,\varepsilon}^{U,u}(v),$$

and by applying this formula we obtain the following fact.

Theorem 7.1 *In the same notations as above for every subset of vertices* $U = \{u\}$ *and every* $k \in \mathbb{N}, \varepsilon > 0$, *there exists a set of weights* $\theta_u = \theta_u(U, k, \varepsilon), u \in U$, *such that for every spline* $Y^{U,y}_{k,\varepsilon}$ *that takes values* $Y^{U,y}_{k,\varepsilon}(u) = y_u$, $u \in U$, *the following exact formula holds:*

$$\sum_{v \in V} Y^{U,y}_{k,\varepsilon}(v) = \sum_{u \in U} y_u \theta_u. \tag{7.2}$$

For a subset $S \subset V$ (finite or infinite) the notation $L_2(S)$ will denote the space of all functions from $L_2(G)$ with support in S:

$$L_2(S) = \{\varphi \in L_2(G),\ \varphi(v) = 0,\ v \in V \backslash S\}.$$

Recall that we say a set of vertices $S \subset V$ is a Λ-set if for any $\varphi \in L_2(S)$ it admits a Poincaré inequality with a constant $\Lambda = \Lambda(S) > 0$,

$$\|\varphi\| \leq \Lambda \|\mathcal{L}\varphi\|,\ \varphi \in L_2(S). \tag{7.3}$$

The infimum of all $\Lambda > 0$ for which S is a Λ-set will be called the Poincaré constant of the set S and denoted by $\Lambda(S)$.

The following Theorem gives a quadrature rule that allows to compute the integral $\sum_{v \in V} f(v)$ by using only values of f on a smaller set U.

Theorem 7.2 *For every set of vertices* $U \subset V$ *for which* $S = V \setminus U$ *is a* Λ*-set and for any* $\varepsilon > 0$, $k = 2^l$, $l \in \mathbb{N}$, *there exist weights* $\theta_u = \theta_u(U, k, \varepsilon)$ *such that for every function* $f \in L_2(G)$,

$$\left| \sum_{v \in V} f(v) - \sum_{u \in U} f(u)\theta_u \right| \leq 2\sqrt{|S|}\Lambda^k \left\| (\varepsilon I + \mathcal{L})^k f \right\|. \tag{7.4}$$

Proof. If $f \in L_2(G)$ and $Y^{U,f}_{k,\varepsilon}$ is a variational spline which interpolates f on a set $U = V \setminus S$ then

$$\left| \sum_{v \in V} f(v) - \sum_{v \in V} Y^{U,f}_{k,\varepsilon}(v) \right| \leq \sum_{v \in S} \left| f(v) - Y^{U,f}_{k,\varepsilon}(v) \right| \leq \sqrt{|S|} \left\| f - Y^{U,f}_{k,\varepsilon} \right\|. \tag{7.5}$$

Since S is a Λ- set we have

$$\|f - Y_{k,\varepsilon}^{U,f}\| \leq \Lambda \|\mathcal{L}\left(f - Y_{k,\varepsilon}^{U,f}\right)\|. \tag{7.6}$$

For any $g \in L_2(G)$ the following inequality holds true:

$$\|\mathcal{L}g\| \leq \|(\varepsilon I + \mathcal{L})g\|. \tag{7.7}$$

Thus one obtains the inequality

$$\left\|f - Y_{k,\varepsilon}^{U,f}\right\| \leq \Lambda \left\|(\varepsilon I + \mathcal{L})\left(f - Y_{k,\varepsilon}^{U,f}\right)\right\|. \tag{7.8}$$

We apply Lemma 5.1 with $A = \varepsilon I + \mathcal{L}$, $a = \Lambda$ and $\varphi = f - Y_{k,\varepsilon}^{U,f}$. It gives the inequality

$$\left\|f - Y_{k,\varepsilon}^{U,f}\right\| \leq \Lambda^k \left\|(\varepsilon I + \mathcal{L})^k \left(f - Y_{k,\varepsilon}^{U,f}\right)\right\| \tag{7.9}$$

for all $k = 2^l, l = 0, 1, 2, ...$ Using the minimization property of $Y_{k,\varepsilon}^{U,f}$ we obtain

$$\left\|f - Y_{k,\varepsilon}^{U,f}\right\| \leq 2\Lambda^k \left\|(\varepsilon I + \mathcal{L})^k f\right\|, k = 2^l, l \in \mathbb{N}.$$

Together with (7.5) it gives

$$\left|\sum_{v \in V} f(v) - \sum_{v \in V} Y_{k,\varepsilon}^{U,f}(v)\right| \leq 2\sqrt{|S|}\Lambda^k \|(\varepsilon I + \mathcal{L})^k f\|, \; k = 2^l, l \in \mathbb{N}. \tag{7.10}$$

By applying the Theorem 7.1 we finish the proof. ■

It is worth noting that the above formulas are optimal in the sense described below.

Definition 7.1 *For the given $U \subset V$, $f \in L_2(G), k \in \mathbb{N}, \varepsilon \geq 0, K > 0$, the notation $Q(U, f, k, \varepsilon, K)$ will be used for a set of all functions h in $L_2(G)$ such that*

1) $h(u) = f(u), u \in U$,

and

2) $\left\|(\varepsilon I + \mathcal{L})^k h\right\| \leq K$.

It is easy to verify that every set $Q(U, f, k, \varepsilon, K)$ is convex, bounded, and closed. It implies that the set of all integrals of functions in $Q(U, f, k, \varepsilon, K)$ is an interval, i. e.,

$$[a,\ b] = \left\{ \sum_{v \in V} h(v) :\ \ h \in Q(U, f, k, \varepsilon, K) \right\}. \tag{7.11}$$

The optimality result is the following.

Theorem 7.3 *For every set of vertices $U \subset V$ and for any $\varepsilon > 0$, $k = 2^l$, $l \in \mathbb{N}$, if $\theta_u = \theta_u(U,\ k,\ \varepsilon)$ are the same weights that appeared in the previous statements, then for any $g \in Q(U, f, k, \varepsilon, K)$,*

$$\sum_{u \in U} g(u)\theta_u = \frac{a+b}{2}, \tag{7.12}$$

where $[a,\ b]$ is defined in (7.11).

Proof. We are going to show that for a given function f the interpolating spline $Y_{k,\varepsilon}^{U,f}$ is the center of the convex, closed and bounded set $Q(U, f, k, \varepsilon, K)$ for any $K \geq \left\| (\varepsilon I + \mathcal{L})^k Y_{k,\varepsilon}^{U,f} \right\|$. In other words, it is sufficient to show that if

$$Y_{k,\varepsilon}^{U,f} + h \in Q(U, f, k, \varepsilon, K)$$

for some function h, then the function $Y_{k,\varepsilon}^{U,f} - h$ also belongs to the same intersection. Indeed, since h is zero on the set U then according to (7.1) one has

$$\left\langle (\varepsilon I + \mathcal{L})^k Y_{k,\varepsilon}^{U,f}, (\varepsilon I + \mathcal{L})^k h \right\rangle = \left\langle (\varepsilon I + \mathcal{L})^{2k} Y_{k,\varepsilon}^{U,f}, h \right\rangle = 0.$$

But then

$$\left\| (\varepsilon I + \mathcal{L})^k (Y_{k,\varepsilon}^{U,f} + h) \right\| = \left\| (\varepsilon I + \mathcal{L})^k \left(Y_{k,\varepsilon}^{U,f} - h \right) \right\|.$$

In other words,

$$\left\| (\varepsilon I + \mathcal{L})^k \left(Y_{k,\varepsilon}^{U,f} - h \right) \right\| \leq K,$$

and because $Y_{k,\varepsilon}^{U,f} + h$ and $Y_{k,\varepsilon}^{U,f} - h$ take the same values on U the function $Y_{k,\varepsilon}^{U,f} - h$ belongs to $Q(U, f, k, \varepsilon, K)$. From here the Theorem follows. ■

Corollary 7.1 *Fix a function $f \in L_2(G)$ and a set of vertices $U \subset V$ for which $S = V \setminus U$ is a Λ-set. Then for any $\varepsilon > 0$, $k = 2^l$, $l \in \mathbb{N}$, for the same set of weights $\theta_u = \theta_u(U,\, k,\, \varepsilon) \in \mathbb{R}$ that appeared in the previous statements the following inequalities hold for every function $g \in Q(U, f, k, \varepsilon, K)$,*

$$\left| \sum_{v \in V} g(v) - \sum_{u \in U} f(u)\theta_u \right| \leq \sqrt{|S|}\Lambda^k diam Q(U, f, k, \varepsilon, K). \tag{7.13}$$

Proof. Since f and g coincide on U from (7.5) and (7.9) we obtain the inequality

$$\left| \sum_{v \in V} g(v) - \sum_{v \in V} Y_{k,\varepsilon}^{U,f}(v) \right| \leq \sqrt{|S|}\Lambda^k \left\| (\varepsilon I + \mathcal{L})^k \left(f - Y_{k,\varepsilon}^{U,f} \right) \right\|. \tag{7.14}$$

By Theorem 7.3 the following inequality holds

$$\left\| (\varepsilon I + \mathcal{L})^k \left(Y_{k,\varepsilon}^{U,f} - g \right) \right\| \leq \frac{1}{2} diam Q(U, f, k, \varepsilon, K)$$

for any $g \in Q(U, f, k, \varepsilon, K)$. The last two inequalities imply the Corollary.

■

7.3 Approximate quadratures for bandlimited functions

We have seen that the operator (matrix) $\mathcal{L}$ is symmetric and positive definite. Let $\mathbf{E}_\omega(\mathcal{L})$ be the span of eigenvectors of $\mathcal{L}$ whose corresponding eigenvalues are $\leq \omega$. The invariant subspace $\mathbf{E}_\omega(\mathcal{L})$ is the space of all vectors in $L_2(G)$ on which $\mathcal{L}$ has norm ω. In other words, f belongs to $\mathbf{E}_\omega(\mathcal{L})$ if and only if the following Bernstein-type inequality holds:

$$\|\mathcal{L}^s f\| \leq \omega^s \|f\|,\ s \geq 0. \tag{7.15}$$

The Bernstein inequality eq.(7.15), the Lemma 5.1, and Theorem 7.2 imply the following result.

Corollary 7.2 *For every set of vertices $U \subset V$ for which $S = V \setminus U$ is a Λ-set and for any $\varepsilon > 0$, $k = 2^l$, $l \in \mathbb{N}$, there exist weights $\theta_u = \theta_u(U, k, \varepsilon) \in \mathbb{R}$ such that for every function $f \in \boldsymbol{E}_\omega(\mathcal{L})$, the following inequality holds:*

$$\left| \sum_{v \in V} f(v) - \sum_{u \in U} f(u)\theta_u \right| \leq 2\gamma^k \sqrt{|S|}\, \|f\|\,, \tag{7.16}$$

where $\gamma = \Lambda(\omega + \varepsilon)$, $k = 2^l$, $l \in \mathbb{N}$.

If, in addition, the following condition holds

$$0 < \omega < \frac{1}{\Lambda} - \varepsilon$$

and $f \in \mathbf{E}_\omega(\mathcal{L})$, then this Corollary implies the following Theorem.

Theorem 7.4 *If U is a subset of vertices for which $S = V \setminus U$ is a Λ-set then for any $0 < \varepsilon < 1/\Lambda$, $k = 2^l$, $l \in \mathbb{N}$, there exist weights $\theta_u = \theta_u(U, k, \varepsilon) \in \mathbb{R}$ such that for every function $f \in \boldsymbol{E}_\omega(\mathcal{L})$, where*

$$0 < \omega < \frac{1}{\Lambda} - \varepsilon,$$

the following relation holds

$$\left| \sum_{v \in V} f(v) - \sum_{u \in U} f(u)\theta_u \right| \to 0, \tag{7.17}$$

when $k = 2^l \to \infty$.

Example 1.

Consider the cycle graph C_{1000} of 1000 vertices. The normalized Laplace operator $\mathcal{L}$ has one thousand eigenvalues which are given by the formula $\lambda_k = 2 - 2\cos\frac{2\pi k}{1000}$, $k = 0, 1, ..., 999$.

It is easy to verify that every single vertex in C_{1000} is a $\Lambda = \frac{1}{\sqrt{6}}$-set. It is also easy to understand that if closures of two vertices do not intersect i. e.,

$$(v_j \cup \partial v_j) \cap (v_i \cup \partial v_i) = \emptyset, \;\; v_j, v_i \in C_{1000},$$

then their union $v_j \cup v_i$ is also a $\Lambda = \frac{1}{\sqrt{6}}$-set. It implies, that one can remove from C_{1000} every third vertex and on the remaining set of 670 the formula (7.17) will be true for the span of about 290 first eigenfunctions of $\mathcal{L}$.

Example 2.
One can show that if $S = \{v_1, v_2, ..., v_N\}$ consists of $|S|$ successive vertices of the graph C_{1000} then it is a Λ-set with

$$\Lambda = \frac{1}{2}\left(\sin\frac{\pi}{2|S|+2}\right)^{-2}.$$

It implies for example that on a set of 100 uniformly distributed vertices of C_{1000} the formula (7.17) will be true for every function in the span of about 40 first eigenfunctions of $\mathcal{L}$.

7.4 Another set of exact and approximate quadratures for bandlimited functions

We introduce another set of quadratures which are exact on some sets of bandlimited functions.

Theorem 7.5 *If U is a subset of vertices for which $S = V \setminus U$ is a Λ-set then there exist weights $\sigma_u = \sigma_u(U) \in \mathbb{R}$, $u \in U$, such that for every function $f \in \boldsymbol{E}_\omega(\mathcal{L})$, where*

$$0 < \omega < \frac{1}{\Lambda},$$

the following exact formula holds

$$\sum_{v\in V} f(v) = \sum_{u\in U} f(u)\sigma_u,\ U = V \setminus S. \tag{7.18}$$

Proof. First, we show that the set U is a uniqueness set for the space $\mathbf{E}_\omega(\mathcal{L})$, i. e., for any two functions from $\mathbf{E}_\omega(\mathcal{L})$ the fact that they coincide on U implies that they coincide on V.

If $f, g \in \mathbf{E}_\omega(\mathcal{L})$ then $f - g \in \mathbf{E}_\omega(\mathcal{L})$ and according to the inequality (7.15) the following holds true:

$$\|\mathcal{L}(f-g)\| \le \omega \|f-g\|. \tag{7.19}$$

If f and g coincide on $U = V \backslash S$ then $f - g$ belongs to $L_2(S)$ and since S is a Λ-set then we will have

$$\|f-g\| \le \Lambda \|\mathcal{L}(f-g)\|, \;\; f-g \in L_2(S).$$

Thus, if $f - g$ is not zero and $\omega < 1/\Lambda$, we have the following inequalities

$$\|f-g\| \le \Lambda \|\mathcal{L}(f-g)\| \le \Lambda\omega \|f-g\| < \|f-g\|, \tag{7.20}$$

which contradict to the assumption that $f - g$ is not identically zero. Thus, the set U is a uniqueness set for the space $\mathbf{E}_\omega(\mathcal{L})$.

It implies that there exists a constant $C = C(U, \omega)$ for which the following Plancherel-Polya inequalities hold true

$$\left(\sum_{u \in U} |f(u)|^2 \right)^{1/2} \le \|f\| \le C \left(\sum_{u \in U} |f(u)|^2 \right)^{1/2} \tag{7.21}$$

for all $f \in \mathbf{E}_\omega(\mathcal{L})$. Indeed, the functional

$$|||f||| = \left(\sum_{u \in U} |f(u)|^2 \right)^{1/2}$$

defines another norm on $\mathbf{E}_\omega(\mathcal{L})$ because the condition $|||f||| = 0, f \in \mathbf{E}_\omega(\mathcal{L})$, implies that f is identically zero on the entire graph. Since in finite-dimensional situation any two norms are equivalent we obtain existence of a constant C for which (7.21) holds true.

Let $\delta_v \in L_2(G)$ be a Dirac measure supported at a vertex $v \in V$. The notation ϑ_v will be used for a function which is orthogonal projection of the function

$$\frac{1}{\sqrt{d(v)}} \delta_v$$

on the subspace $\mathbf{E}_\omega(\mathcal{L})$. If $\varphi_0, \varphi_1, ..., \varphi_{j(\omega)}$ are orthonormal eigenfunctions of $\mathcal{L}$ which constitute an orthonormal basis in $\mathbf{E}_\omega(\mathcal{L})$, then the explicit formula for ϑ_v is

$$\vartheta_v = \sum_{j=0}^{j(\omega)} \varphi_j(v)\varphi_j. \tag{7.22}$$

In these notations the Plancherel-Polya inequalities (7.21) can be written in the form

$$\sum_{u\in U} |\langle f, \vartheta_u\rangle|^2 \le \|f\|^2 \le C^2 \sum_{u\in U} |\langle f, \vartheta_u\rangle|^2, \tag{7.23}$$

where $f, \vartheta_u \in \mathbf{E}_\omega(\mathcal{L})$ and $\langle f, \vartheta_u\rangle$ is the inner product in $L_2(G)$. These inequalities mean that if U is a uniqueness set for the subspace $\mathbf{E}_\omega(\mathcal{L})$, then the functions $\{\vartheta_u\}_{u\in U}$ form a frame in the subspace $\mathbf{E}_\omega(\mathcal{L})$ and the tightness of this frame is $1/C^2$. This fact implies that there exists a frame of functions $\{\Theta_u\}_{u\in U}$ in the space $\mathbf{E}_\omega(\mathcal{L})$ such that the following reconstruction formula holds true for all $f \in \mathbf{E}_\omega(\mathcal{L})$:

$$f(v) = \sum_{u\in U} f(u)\Theta_u(v), v \in V. \tag{7.24}$$

By setting $\sigma_u = \sum_{v\in V} \Theta_u(v)$ one obtains (7.18).

Unfortunately this approach does not give any information about the constant C in (7.23) and it makes realization of the Theorem 7.5 problematic. We are going to utilize another approach to the Plancherel-Polya-type inequality which was developed in [31] and which produces an explicit constant.

Definition 7.2 *The weighted gradient of a function f on $V(G)$ is defined by*

$$\|\nabla_w f\| = \left(\sum_{u,v\in V(G)} \frac{1}{2}|f(u) - f(v)|^2 w(u,v) \right)^{1/2}.$$

The set of all $f : G \to \mathbb{C}$ for which the weighted gradient is finite will be denoted as $\mathcal{D}^2(\nabla_w)$.

The weighted Laplace operator $L_w : L_2(G) \to L_2(G)$ is introduced via

$$(L_w f)(v) = \sum_{u\in V(G)} (f(v) - f(u))w(v,u)\ . \tag{7.25}$$

This graph Laplacian is a well-studied object; it is known to be a positive-semidefinite self-adjoint *bounded* operator.

What is really important for us is that for the non-negative square root $L_w^{1/2}$ one has the equality

$$\|L_w^{1/2} f\|_2 = \|\nabla_w f\|_2 \tag{7.26}$$

for all $f \in \mathcal{D}^2(\nabla_w)$.

Lemma 7.1 *For all $f \in L_2(G)$ contained in the domain of $L_w^{1/2}$, we have*

$$\|L_w^{1/2} f\|_2^2 = \|\nabla_w f\|_2^2 \ . \tag{7.27}$$

For $f \in PW_\omega(G)$, this implies

$$\|\nabla_w f\|_2 = \|L_w^{1/2} f\|_2 \leq \sqrt{\omega}\|f\|_2. \tag{7.28}$$

Proof. Let, as above, $d(u) = w_{V(G)}(u)$, the weighted degree of u. Then we obtain

$$\begin{aligned}
\langle f, L_w f\rangle &= \sum_{u\in V(G)} f(u)\overline{\left(\sum_{v\in V(G)} (f(u)-f(v))\, w(u,v)\right)} \\
&= \sum_{u\in V(G)} \left(|f(u)|^2 d(u) - \sum_{v\in V(G)} f(u)\overline{f(v)} w(u,v)\right) \ .
\end{aligned}$$

In the same way

$$\begin{aligned}
\langle f, L_w f\rangle &= \langle L_w f, f\rangle \\
&= \sum_{u\in V(G)} \left(|f(u)|^2 d(u) - \sum_{v\in V(G)} \overline{f(u)} f(v) w(u,v)\right) \ .
\end{aligned}$$

Averaging these equations yields

$$\begin{aligned}
\langle f, L_w f\rangle &= \sum_{u\in V(G)} \left(|f(u)|^2 d(u) - \mathrm{Re} \sum_{v\in V(G)} f(u)\overline{f(v)} w(u,v)\right) \\
&= \frac{1}{2} \sum_{u,v\in V(G)} |f(u)|^2 w(u,v) + |f(v)|^2 w(u,v) - 2\mathrm{Re} f(u)\overline{f(v)} w(u,v) \\
&= \sum_{u,v\in V(G)} \frac{1}{2}|f(v)-f(u)|^2 w(u,v) = \|\nabla_w f\|_2^2 \ .
\end{aligned}$$

Now the first equality follows by taking the square root of L_w (note that by spectral theory, f is also in the domain of $L_w^{1/2}$), and (7.28) is an obvious consequence. ■

We recall that for any U which is a subset of vertices of G we introduced (see section 5.3) the following operator:

$$cl^0(U) = S,\ \ cl(U) = U \cup \partial U, cl^m(U) = cl\left(cl^{m-1}(U)\right), m \in \mathbb{N}, U \subset V. \quad (7.29)$$

We will use the following notion of the relative degree.

Definition 7.3 *For a vertex $v \in cl^m(U)$ we introduce the relative degree $d_m(v)$ as the number of vertices in the boundary $\partial\left(cl^m(U)\right)$ which are adjacent to v:*

$$d_m(v) = card\left\{w \in \partial\left(cl^m(U)\right) : w \sim v\right\}.$$

For any $U \subset V$ we introduce the following notation

$$D_m = D_m(U) = \sup_{v \in cl^m(U)} d_m(v).$$

Definition 7.4 *For a vertex $v \in \partial\left(cl^m(U)\right)$ we introduce the quantity $k_m(v)$ as the number of vertices in the set $cl^m(U)$ which are adjacent to v:*

$$k_m(v) = card\left\{w \in cl^m(U) : w \sim v\right\}.$$

For any $U \subset V$ we introduce the following notation

$$K_m = K_m(U) = \inf_{v \in \partial(cl^m(U))} k_m(v).$$

For a given set $U \subset V$ and a fixed $n \in \mathbb{N}$ consider a sequence of closures

$$U, cl(U), ..., cl^n(U), n \in \mathbb{N}.$$

Let $U \subset V(G)$ be such that $\partial(cl^{n-1}(U)) \neq \emptyset$, the property $cl^n(U) = V(G)$ is not needed in this context. We let $U_m = \partial(cl^{m-1}(U))$, for $m \geq 1$, and $U_0 = U$.

For $0 \leq m < n$ and $v \in U_m$, we let $\hat{k}_m(v) = \mathrm{card}(\{w \in U_{m+1} : w \sim v\})$, and let

$$\widehat{K}_m = \inf_{v \in U_m} \hat{k}_m(v) \ .$$

By definition of $\widehat{K}_m$, there exist mappings $\hat{v}_1, \ldots, \hat{v}_{\widehat{K}_m} : U_m \to U_{m+1}$ such that for all $u \in U_m$, the map $j \mapsto \hat{v}_j(u)$ is injective. For $v \in U_{m+1}$, we define $\hat{d}_m(v) = \mathrm{card}(\{(j,u) : v = \hat{v}_j(u)\})$, and let

$$\widehat{D}_m = \sup_{v \in U_{m+1}} \hat{d}_m(v) \ .$$

We stress that these quantities may depend on the choice of the $\hat{v}_j$.

Assume that $U \subset V(G)$ is a subset with $cl^n(U) = V(G)$, as well as $\partial(cl^{n-1}(U)) \neq \emptyset$, and define

$$\delta_U = \left(\sum_{m=1}^{n} \left(\sum_{k=1}^{m} \frac{1}{K_{k-1}} \left(\prod_{i=k}^{m-1} \frac{D_i}{K_i} \right) \right) \right)^{1/2}$$

and

$$a_U = \left(\sum_{m=0}^{n} \prod_{j=0}^{m-1} \frac{D_j}{K_j} \right)^{1/2} .$$

Assume also that, for given $n' \leq n$, the constants

$$\hat{\delta}_U = \left(\sum_{m=1}^{n'} \left(\sum_{k=0}^{m-1} \frac{1}{\widehat{K}_k} \left(\prod_{i=k}^{m-1} \frac{\widehat{K}_i}{\widehat{D}_i} \right) \right) \right)^{1/2}$$

and

$$\hat{a}_U = \left(\sum_{m=0}^{n'} \prod_{j=0}^{m-1} \frac{\widehat{K}_j}{\widehat{D}_j} \right)^{1/2} ,$$

are well-defined, i.e., $\widehat{K}_m, \widehat{D}_m > 0$, for $0 \leq m < n'$.

Note, that if f belongs to the space $\mathbf{E}_\omega(\mathcal{L})$ then the Bernstein inequality gives

$$\|\nabla f\| = \sqrt{2} \|\mathcal{L}^{1/2} f\| \leq \sqrt{2\omega} \|f\|. \tag{7.30}$$

The following theorem follows from results in [32].

Theorem 7.6 *Assume that $U \subset V(G)$ is a subset with*

$$cl^n(U) = V(G), \quad \partial(cl^{n-1}(U)) \neq \emptyset. \tag{7.31}$$

Then, if the inequality

$$\delta_U \sqrt{2\omega} < 1, \quad \omega > 0, \tag{7.32}$$

is satisfied, then the following Plancherel-Polya-type equivalence holds for all $f \in \boldsymbol{E}_\omega(\mathcal{L})$:

$$\frac{1 - \delta_U \sqrt{2\omega}}{a_U} \|f\| \leq \|f|_U\| \leq \frac{1 + \hat{\delta}_U \sqrt{2\omega}}{\hat{a}_U} \|f\|. \tag{7.33}$$

Using this result we prove existence of exact quadratures on spaces of bandlimited functions.

Theorem 7.7 *If U is a subset of vertices for which condition (7.31) is satisfied then there exists a set of weights $\mu_u \in \mathbb{R}$, $u \in U$, such that for any $f \in \boldsymbol{E}_\omega(\mathcal{L})$, where ω satisfies (7.32) the following exact formula holds:*

$$\sum_{v \in V} f(v) = \sum_{u \in U} f(u) \mu_u. \tag{7.34}$$

Proof. The previous Theorem shows that U is a uniqueness set for the space $\mathbf{E}_\omega(\mathcal{L})$, which means that every f in $\mathbf{E}_\omega(\mathcal{L})$ is uniquely determined by its values on U.

Let us denote by θ_v, where $v \in U$, the orthogonal projection of the Dirac measure δ_v, $v \in U$, onto the space $\mathbf{E}_\omega(\mathcal{L})$. Since for functions in $\mathbf{E}_\omega(\mathcal{L})$ one has $f(v) = \langle f, \theta_v \rangle$, $v \in U$, the inequality (7.33) takes the form of a frame inequality in the Hilbert space $H = \mathbf{E}_\omega(\mathcal{L})$

$$\left(\frac{1 - \epsilon \delta_U}{a_U} \right)^2 \|f\|^2 \leq \sum_{v \in U} |\langle f, \theta_v \rangle|^2 \leq \left(\frac{1 + \epsilon \hat{\delta}_U}{\hat{a}_U} \right)^2 \|f\|^2, \ \epsilon = \sqrt{2\omega}, \tag{7.35}$$

for all $f \in \mathbf{E}_\omega(\mathcal{L})$. According to the general theory of Hilbert frames [89] the last inequality implies that there exists a dual frame (which is not unique in general) $\{\Theta_v\}$, $v \in U$, $\Theta_v \in \mathbf{E}_\omega(\mathcal{L})$, in the space $\mathbf{E}_\omega(\mathcal{L})$ such that for all $f \in \mathbf{E}_\omega(\mathcal{L})$ the following reconstruction formula holds:

$$f = \sum_{v \in U} f(v) \Theta_v. \tag{7.36}$$

By setting $\sum_{v\in V}\Theta_v(u)=\mu_u$ we obtain (7.34). ■

To be more specific we consider the case when

$$cl\,U = U\cup\partial U = V(G). \tag{7.37}$$

In this case

$$d_0(v) = card\left\{w\in bU : w\sim v\right\},\quad D_0 = D_0(U) = \sup_{v\in U} d_0(v),$$

and

$$k_0(v) = card\left\{w\in U : w\sim v\right\},\quad K_0 = K_0(U) = \inf_{v\in bU} k_0(v).$$

It is easy to see that

$$a_U = \left(1+\frac{D_0}{K_0}\right)^{1/2},\quad \delta_U = \frac{1}{K_0^{1/2}}.$$

Thus, we have

$$\|f\| \le \left(1+\frac{D_0}{K_0}\right)^{1/2}\|f_0\| + \frac{1}{K_0^{1/2}}\|\nabla f\|.$$

By applying (7.15) along with the assumption

$$\omega < \frac{K_0}{2}, \tag{7.38}$$

we obtain the following estimate

$$\|f\| \le \left(1-\sqrt{\frac{2\omega}{K_0}}\right)\left(1+\frac{D_0}{K_0}\right)^{1/2}\|f_0\|,\quad f_0 = f|_U. \tag{7.39}$$

On the other hand,

$$\hat{a}_U = \left(1+\frac{\hat{K}_0}{\hat{D}_0}\right)^{1/2},\quad \hat{\delta}_U = \frac{1}{\hat{D}_0^{1/2}}.$$

This yields the norm estimate

$$\|f\| + \frac{1}{\hat{D}_0^{1/2}}\|\nabla f\| \ge \left(1+\frac{\hat{K}_0}{\hat{D}_0}\right)^{1/2}\|f_0\|,\quad f_0 = f|_U.$$

If (7.15) holds, then

$$\left(1+\frac{\hat{K}_0}{\hat{D}_0}\right)^{1/2}\|f_0\| \le \|f\|_2 + \frac{1}{\hat{D}_0^{1/2}}\|\nabla f\| \le \left(1+\sqrt{\frac{2\omega}{\hat{D}_0}}\right)\|f\|. \tag{7.40}$$

After all, for functions f in $\mathbf{E}_\omega(\mathcal{L})$ with $\omega < K_0/2$, we obtain the following frame inequality:

$$A\|f\|^2 \le \sum_{v\in U} |< f, \theta_v >|^2 \le B\|f\|^2, \quad f_0 = f|_U, \tag{7.41}$$

where

$$A = \frac{\left(1-\sqrt{\frac{2\omega}{K_0}}\right)^2}{1+\frac{D_0}{K_0}}, \quad B = \frac{\left(1+\sqrt{\frac{2\omega}{\widehat{D_0}}}\right)^2}{1+\frac{\widehat{K_0}}{\widehat{D_0}}}. \tag{7.42}$$

It shows that if the condition $\omega < K_0/2,\ K_0 = K_0(U)$ is satisfied, then the set U is a sampling set for the space $\mathbf{E}_\omega(\mathcal{L})$ and a reconstruction formula (7.36) holds.

However, it is not easy to find a dual frame $\{\Theta_v\}$, $v \in U$. For this reason we are going to adopt the frame algorithm for reconstruction of functions in $\mathbf{E}_\omega(\mathcal{L})$ from the set U.

Let $\{e_j\}$ be a frame in a Hilbert H space with frames bounds A, B, i. e.,

$$A\|f\|_H^2 \le \sum_j |\langle f, e_j\rangle|^2 \le B\|f\|_H^2, \ \ f \in H.$$

Given a relaxation parameter $0 < \nu < \frac{2}{B}$, set $\eta = \max\{|1-\nu A|, |1-\nu B|\} < 1$. Let $f_0 = 0$ and define recursively

$$f_n = f_{n-1} + \nu\Phi(f - f_{n-1}), \tag{7.43}$$

where Φ is the frame operator which is defined on H by the formula

$$\Phi f = \sum_j \langle f, e_j\rangle e_j.$$

In particular, $f_1 = \nu\Phi f = \nu\sum_j \langle f, e_j\rangle e_j$. Then $\lim_{n\to\infty} f_n = f$ with a geometric rate of convergence, that is,

$$\|f - f_n\|_H \le \eta^n \|f\|_H. \tag{7.44}$$

Note, that for the choice $\nu = \frac{2}{A+B}$ the convergence factor is

$$\eta = \frac{B-A}{A+B}. \tag{7.45}$$

In our situation the frame operator Φ takes the following form

$$\Phi f = \sum_{v \in U} \langle f, \theta_v \rangle \, \theta_v, \;\; f \in \mathbf{E}_\omega(\mathcal{L}), \tag{7.46}$$

where

$$\langle f, \theta_v \rangle = f(v), \;\; f \in \mathbf{E}_\omega(\mathcal{L}), \; v \in U.$$

Thus the recurrence sequence (7.43) takes the form $f_0 = 0$, and

$$f_n = f_{n-1} + \nu \sum_{v \in U} (f - f_{n-1})(v) \theta_v. \tag{7.47}$$

We are ready to state the following fact which provides a reconstruction method for functions in $\mathbf{E}_\omega(\mathcal{L})$ from their values on U.

Theorem 7.8 *Under the assumptions (7.37), (7.38) for $f \in \boldsymbol{E}_\omega(\mathcal{L})$ the following inequality holds for all natural n*

$$\|f - f_n\| \leq \eta^n \|f\|, \tag{7.48}$$

where the convergence factor η is given by (7.45), (7.42).

By setting

$$\nu_u = \sum_{v \in V} \theta_u(v), \; u \in U,$$

we obtain the recurrence sequence for corresponding integrals

$$\sum_{v \in V} f_0(v) = 0,$$

$$\sum_{v \in V} f_1(v) = \lambda \sum_{u \in U} f(u) \nu_u,$$

$$\sum_{v \in V} f_n(v) = \sum_{v \in V} f_{n-1}(v) + \lambda \sum_{u \in U} (f - f_{n-1})(u) \nu_u. \tag{7.49}$$

Thus, the integral $\sum_{v\in V} f_n(v)$ for every f_n is expressed in terms of weights $\{\nu_u\}$, $u \in U$, and values of f on the uniqueness set U.

We are ready to state the following fact which provides an approximate quadrature for functions in $\mathbf{E}_\omega(\mathcal{L})$. It should be noted that the effectiveness of this result depends on one's ability to implement the frame operator Φ.

Theorem 7.9 *Under the assumptions (7.37), (7.38) for $f \in \boldsymbol{E}_\omega(\mathcal{L})$ and $\sum_{v\in V} f_n(v)$ defined in* (7.49) *we have*

$$\left|\sum_{v\in V} f(v) - \sum_{v\in V} f_n(v)\right| \le \eta^n \sqrt{|V|}\|f\|,$$

where the convergence factor η is given in (7.45), (7.42).

The proof is obvious:

$$\left|\sum_{v\in V} f(v) - \sum_{v\in V} f_n(v)\right| \le \sum_{v\in V} |f(v) - f_n(v| \le \sqrt{|V|}\|f - f_n\| \le \eta^n \sqrt{|V|}\|f\|.$$

CHAPTER 8

SPECTRAL GRAPH DRAWING USING SPLINE SPACES

Graphs are abstract models of a set of objects and some kind of pairwise relationship that exists between some pairs of the objects. The objects are called nodes and the relationships are called edges or links. That is, graphs are used to model connections or relations between things (data points), where vertices represent data points and edges represent the connections or relations. Graphs provide a flexible model for representing data in many domains such as networks, computer vision, and high dimensional data. The data on the graphs can be visualized as a finite collection of samples, leading to a graph signal, which can be defined as the information attached to each node of the graph.

The strength of the relation among the objects can be described by the weight of edges, while the amount of information on each node can be represented by a real-valued function defined on the vertex set $V = V(G)$. Given a set of data points, the goal of graph drawing is to try to draw graphs in such a way that they convey the most information possible. It is a standard means for visualizing relational information. To visualize infor-

mation expressed as graphs, many researchers have been working on graph drawing and many successful approaches and methods have been developed [46, 47, 48, 49, 50, 51, 52, 53, 57, 56]. The graph drawing problem is to compute an aesthetically good quality layout of vertices and edges so that it is easy to grasp visually the inherent structure of the graph. Depending on the aesthetic criteria of interest, various approaches have been developed. A popular approach is to define an energy function and to iteratively compute a local minimum of the energy function [46]. The positions of the vertices at the local minimum produce the final layout of the graph. The term spectral graph drawing refers to an approach that produces a final layout of a graph using the spectral decomposition of some matrix derived from the vertex and edge sets of the graph.

As the name indicates, spectral layout denotes the use of eigenvectors of graph related matrices such as the adjacency or Laplacian matrix as coordinate vectors. The idea of graph drawing by using eigenvectors goes back to the work of Kenneth Hall in 1970 [50]. Hall computed the layout of a graph using some generalized eigenvectors of the related Laplacian matrix. Many developments have been made since then.

8.1 Graph Representation

The idea of the graph Laplacian matrix being the discrete version of the continuous Laplacian operator Δ was discussed in recent literature (see for example [99], sec 3.3). In the continuous case, it is well known that the energy functional (defined below) is minimized over all functions whose Laplacian is zero. To compare the energy functional defined in the discrete setting we first give the following definition and the next theorem from [99] which will be applicable to the continuous Laplacian operator.

Definition 8.1 *Let U be an open bounded region in $\mathbb{R}^k$. The energy functional*

$E(u)$ for which a function $u : U \to \mathbb{R}$ is

$$E(u) = \frac{1}{2} \int_U \| \bigtriangledown u\|^2 \, dx, \tag{8.1}$$

where $\bigtriangledown u$ is the gradient vector $(u_{x_1}, u_{x-2}, \ldots, u_{x_m})$.

In the following theorem from [99] known as Dirichlet's principle, let $\overline{U}$ and ∂U denote the closure and boundary of the region U, respectively, and let $C^2(U)$ denote the set of functions that are continuously twice differentiable on U.

Theorem 8.1 *Let U be an open bounded region in $\mathbb{R}^m$, let $u : U \to \mathbb{R}$, and let $A = \{w \in C^2(\overline{U}) \ : \ w|_{\partial U} = 0\}$. Then $\Delta u = 0$ on $\overline{U}$ if and only if*

$$E(u) = \min_{w \in A} E(w).$$

We now turn our attention to graphs. Our focus will be drawing graphs in $\mathbb{R}^m$ where the energy function is minimized.

Definition 8.2 *Let G be a graph with vertex set V. A representation of G in $\mathbb{R}^m$ is a mapping $\rho : V \to \mathbb{R}^m$.*

Intuitively, we think of a representation as the positions of the vertices in an m-dimensional drawing of a graph. In other words, ρ represents the position of the vertices when we draw G in $\mathbb{R}^m$. It is often convenient to represent a representation as a representation matrix X where row i of X corresponds to $\rho(i)$, the representation of vertex i. So we regard the vectors $\rho(u)$ as row vectors, and thus we may represent ρ by the $|V(G)| \times m$ matrix X with the images of the vertices of G as its rows.

Regarding the vertex set of a graph as a subset $\mathbb{R}^m$ in a similar way we regard U as a subset of $\mathbb{R}^m$, observe that ρ is a discrete function from a subset of $\mathbb{R}^m$ to $\mathbb{R}$ as u was a continuous such function. Therefore, we can define the discrete version of energy function with respect to a domain. However, in the discrete version, the domain is a vertex set of a graph rather than an open bounded set $U \subset \mathbb{R}^m$.

Adapting the notion of energy to graph theory, we will see that the eigenvalues and eigenvectors of the Laplacian matrix will be vital in drawing the graph in $\mathbb{R}^m$ so that the energy of the graph is minimized.

Definition 8.3 *Let G be a graph with vertex set V represented in $\mathbb{R}^m$ with representation $\rho : V \to \mathbb{R}$. The energy $\mathcal{E}(\rho)$ of the representation ρ with respect to the graph G is the quantity*

$$\mathcal{E}(\rho) = \frac{1}{2} \sum_{\substack{v \in V \\ u \sim v}} \|\rho(u) - \rho(v)\|^2. \tag{8.2}$$

Given an edge ij of a graph, observe in eq.(8.2) that each edge is being accounted for twice, once when $i = u$ and $j = v$ and again when $i = v$ and $j = u$. Hence we can rewrite eq.(8.2) as

$$\mathcal{E}(\rho) = \sum_{uv \in E(G)} \|\rho(u) - \rho(v)\|^2. \tag{8.3}$$

Definition 8.4 *A representation ρ of a graph G with vertex set V is balanced if*

$$\sum_{u \in V(G)} \rho(u) = 0. \tag{8.4}$$

Thus if ρ is represented by X, then ρ is balanced if and only if $\mathbf{1}^T X = 0$.

Given a graph G on n vertices and a positive integer $m \leq n - 1$, our goal is to find a representation ρ that minimizes the energy. To this end, observe that if G is drawn in $\mathbb{R}^m$, we can translate the graph so that it is balanced without changing the quantities $\|\rho(u) - \rho(v)\|^2$ or without losing any information. Hence from this point on, we can assume that a graph G is drawn so that its representation ρ is balanced. If the columns of the matrix X are not linearly independent, then the image of G is contained in a proper subspace of $\mathbb{R}^m$ and ρ is just a lower dimensional representation embedded in $\mathbb{R}^m$. Any maximal linearly independent subset of the columns of X would suffice to determine all the properties of the representation. Therefore, we will assume that the columns of X are linearly independent.

We can imagine building a physical model of G by placing the vertices in the positions specified by ρ and connecting adjacent vertices by identical springs (i.e., all edges have the same weight, or equivalently the graph is unweighted). It is natural to consider a representation to be better if it requires to be less extended. This means that a good drawing of graphs correspond to representations with low energy. Of course, the representation with least energy is the one where each vertex is mapped to the zero vector. Thus we need to add further constraints to exclude this.

In the case of weighted graphs, the energy $\mathcal{E}(\rho)$ of a representation ρ of G is defined by

$$\mathcal{E}(\rho) = \sum_{uv \in E(G)} w(u,v)||\rho(u) - \rho(v)||^2, \tag{8.5}$$

where $w(u,v)$ is the weight of the edge uv.

Graph representation in $\mathbb{R}^m$ or an m-dimensional drawing is also called an *m-dimensional layout*. Describing the above discussion in another way, an m-dimensional layout of a graph with n vertices is defined by m vectors, $x^1, \ldots, x^m \in \mathbb{R}^n$, where $x^1(i), \ldots, x^m(i)$ are the coordinates of node i. In other words, an m-dimensional layout of the graph is given by the $n \times m$ matrix

$$X = [x^1 \ldots x^m],$$

where the i-th row provides the embedding coordinates of the i-th vertex. Since we can only visualize graphs in low dimensions (usually in one, two, or three dimensions), for information visualization and most applications, $m = 2$ or 3, though it can be any positive integer less than n in general.

Theorem 8.2 *[99] Let G be a weighted graph with n vertices and let ρ be its representation in $\mathbb{R}^m$ and given by the $n \times m$ matrix X. Then*

$$\mathcal{E}(\rho) = tr\, X^T L X, \tag{8.6}$$

where L is the (weighted) Laplacian of G.

Note that $X^T LT$ is an $m \times m$ symmetric matrix; hence its eigenvalues are real. The sum of the eigenvalues is the trace of the matrix, and hence the energy of the representation is given by the sum of the eigenvalues of $X^T LX$.

The energy of certain representations of a graph G are determined by the eigenvalues of the Laplacian of G. If Y is an invertible $m \times m$ matrix, then the map that sends u to $\rho(u)Y$ is another representation of G. This representation is given by the matrix XY and provides as much information about G as does ρ. From this point of view, the representation is determined by its column space. Therefore we may assume that the columns of X are orthonormal. In this case the matrix X satisfies $X^T X = I_m$, and the representation is called an *orthogonal representation.*

Theorem 8.3 *[99] Let G be a connected weighted graph on n vertices with weighted Laplacian L and Laplacian eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. The minimum energy of a balanced orthogonal representation of G in $\mathbb{R}^m$ equals $\sum_{i=2}^{m+1} \lambda_i$.*

This result provides a nice method for drawing a graph in any number of dimensions. Compute an orthonormal basis of eigenvectors $x_1, \ldots, x_n$ for the Laplacian L and let the columns of X be $x_2, \ldots, x_{m+1}$. Theorem 8.2 implies that this yields an orthogonal balanced representation of minimum energy. However, the representation is not necessarily unique, because it may be the case that $\lambda_{m+1} = \lambda_{m+2}$, in which case there is no reason to choose between x_{m+1} and x_{m+2}.

8.2 Generalized Eigenvectors and Minimum Energy

Definition 8.5 *For a symmetric positive semi-definite matrix A and positive definite matrix B, a vector ν and a scalar λ are called* generalized eigenpairs *of (A, B), if*

$$A\nu = \lambda B\nu. \tag{8.7}$$

In this case we say (ν, λ) is a generalized eigenpair of (A, B).

Theorem 8.4 *All the generalized eigenvalues of (A, B) are real and non-negative, and all the generalized eigenvectors are B-orthogonal.*

Proof: Since B is positive definite, there exists an invertible matrix C such that $B = C^T C$. Therefore,

$$A\nu = \lambda B\nu \implies A\nu = \lambda C^T C\nu \implies (C^{-T}AC^{-1})C\nu = \lambda C\nu.$$

Hence the eigenvalues of $A\nu = \lambda B\nu$ are those of $Mv = \lambda v$, where M is the symmetric matrix $M = C^{-T}AC^{-1}$ and $v = C\nu$. ■

The first few eigenvectors of the Laplacian and the generalized eigenvectors of (L, D) play a significant role in graph drawing (eg. see [46]). They are *optimal* solutions of some constrained minimization problem as explained in the following theorem.

Theorem 8.5 *[46] Given a symmetric matrix $A_{n\times n}$, denote by $\nu^1, \ldots, \nu^n$, its eigenvectors, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then, $\nu^1, \ldots, \nu^p$ $(p < n)$ are optimal solutions of the constrained minimization problem*

$$\begin{aligned} \min_{x^1,\ldots,x^p \in \mathbb{R}^n} \quad & \sum_{k=1}^{p} (x^k)^T A x^k, \\ \text{subject to:} \quad & (x^k)^T x^l = \delta_{kl}, \quad k, l = 1, \ldots, p. \end{aligned} \tag{8.8}$$

The next theorem is a more general form of Theorem 8.5

Theorem 8.6 *Given a symmetric matrix $A_{n\times n}$ and positive definite matrix $B_{n\times n}$, denote by $\nu^1, \ldots, \nu^n$, the generalized eigenvectors of (A, B), with corresponding eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. Then, $\nu^1, \ldots, \nu^p$ are optimal solutions of the constrained minimization problem,*

$$\begin{aligned} \min_{x^1,\ldots,x^p \in \mathbb{R}^n} \quad & \sum_{k=1}^{p} (x^k)^T A x^k, \\ \text{subject to:} \quad & (x^k)^T B x^l = \delta_{kl}, \quad k, l = 1, \ldots, p. \end{aligned} \tag{8.9}$$

Proof: Since B is positive definite, there exists an invertible matrix C such that $B = C^T C$. Substitute $x^k = C^{-1}y^k$ in (8.8), and rewrite the minimization problem as follows:

$$\begin{aligned} \min_{y^1,\ldots,y^p \in \mathbb{R}^n} \quad & \sum_{k=1}^{p} (y^k)^T C^{-T} A C^{-1} y^k, \\ \text{subject to:} \quad & (y^k)^T y^l = \delta_{kl}, \quad k,l = 1,\ldots,p. \end{aligned} \tag{8.10}$$

Let $\hat{y}^1, \ldots, \hat{y}^p$ be the minimizers of (8.10) (and of course this implies $C^{-1}\hat{y}^1, \ldots, C^{-1}\hat{y}^p$ are the minimizers of (8.9).) According to Theorem 8.5, these are the p lowest eigenvectors of the symmetric matrix $C^{-T}AC^{-1}$. Thus, we have $C^{-T}AC^{-1}\hat{y}^k = \lambda_k \hat{y}^k$, $1 \le k \le p$. Now use this equation and substitute back to $\hat{x}^k = C^{-1}\hat{y}^k$ to get

$$C^{-T}A\hat{x}^k = \lambda_k C\hat{x}^k.$$

But this implies $A\hat{x}^k = \lambda_k B\hat{x}^k$, so the minimizers of (8.10) are the lowest generalized eigenvectors of (A, B). ■

An equivalent version of the above theorem and yet most suitable to use is the following:

Theorem 8.7 *Given a symmetric matrix $A_{n\times n}$ and positive definite matrix $B_{n\times n}$, denote by $\nu^1, \ldots, \nu^n$, the generalized eigenvectors of (A, B), with corresponding eigenvalues $\lambda_1 \le \cdots \le \lambda_n$. Then, $\nu^1, \ldots, \nu^p$ are optimal solutions of the constrained minimization problem,*

$$\begin{aligned} \min_{x^1,\ldots,x^p \in \mathbb{R}^n} \quad & \frac{\sum_{k=1}^{p} (x^k)^T A x^k}{\sum_{k=1}^{p} (x^k)^T B x^k}, \\ \textit{subject to:} \quad & (x^1)^T B x^1 = (x^2)^T B x^2 = \cdots = (x^p)^T B x^p, \\ & (x^k)^T B x^l = 0, \quad 1 \le k \ne l \le p. \end{aligned} \tag{8.11}$$

Note that Theorem 8.7 can be reduced to Theorem 8.6 by noting that the minimization problem in Theorem 8.7 is invariant under scaling. Indeed, for

any constant $c \neq 0$,

$$\frac{\sum_{k=1}^{p}(x^k)^T A x^k}{\sum_{k=1}^{p}(x^k)^T B x^k} = \frac{\sum_{k=1}^{p}(cx^k)^T A (cx^k)}{\sum_{k=1}^{p}(cx^k)^T B (cx^k)}.$$

Hence, we can always scale the optimal solution, so that

$$(x^1)^T B x^1 = (x^2)^T B x^2 = \cdots = (x^p)^T B x^p = 1,$$

and reduce the problem to Theorem 8.6

The following theorem is also obtained from the last theorem by imposing the additional restriction that the solution must be B-orthogonal to the lowest generalized eigenvectors. In this case, we simply take the next low generalized eigenvectors.

Theorem 8.8 *Given a symmetric matrix $A_{n\times n}$ and positive definite matrix $B_{n\times n}$, denote by $\nu^1, \ldots, \nu^n$, the generalized eigenvectors of (A, B), with corresponding eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. Then, $\nu^{k+1}, \ldots, \nu^{k+p}$, are an optimal solutions of the constrained minimization problem,*

$$\begin{aligned} \min_{x^1,\ldots,x^p \in \mathbb{R}^n} \quad & \frac{\sum_{i=1}^{p}(x^i)^T A x^i}{\sum_{i=1}^{p}(x^i)^T B x^i}, \\ \textit{subject to:} \quad & (x^1)^T B x^1 = (x^2)^T B x^2 = \cdots = (x^p)^T B x^p, \\ & (x^i)^T B x^j = 0, \quad 1 \leq i \neq j \leq p. \\ & (x^i)^T B \nu^j = 0, \quad i = 1, \ldots, p, \;\; j = 1, \ldots, k. \end{aligned} \tag{8.12}$$

8.3 Spectral Graph Drawing: Eigenprojection Method

Graphs are often used to express the relationship between items. Graph drawing enables visualization of these relationships. The usefulness of the visual representation depends upon whether the drawing is aesthetic. While there are no strict criteria for aesthetic drawing, it is generally agreed that such a drawing has minimal edge crossing and even spacing between vertices.

The problem of graph drawing has been studied extensively in the literature and many approaches have been proposed such as eigenprojection method [46], high dimensional embedding [47], graph drawing using spectral distance embedding [57], and force-directed algorithms [58] are a few to list.

The approach to using eigenvectors to draw graphs was suggested by Kenneth Hall in 1970 [50]. Hall first considered the problem of assigning a real number $x(v)$ to each vertex v so that $(x(v) - x(u))^2$ is small for most edges (v, u). This led him to consider the problem of minimizing (8.5). So as to avoid the degenerate solutions in which every vertex is mapped to zero or any other constant value, he introduced the restriction that x be orthogonal to $\mathbf{1}_n$. As the utility of the embedding does not really depend up on its scale, he suggested the normalization of $||x|| = 1$. According to Theorem 8.5, the solution of the resulting optimization problem is precisely an eigenvector corresponding to the second smallest eigenvalue of the Laplacian. Hall also considered the problem of embedding in $\mathbb{R}^2$. This time he considered the minimization problem

$$\min_{x,y \in \mathbb{R}^n} \sum_{(i,j) \in E} ||(x(i), y(i)) - (x(j), y(j))||^2$$

such that

$$\sum_{(i,j) \in E} (x(i), y(i)) = (0, 0),$$

i.e., the representation is balanced.
However, doing so typically results in the degenerate solution $x = y = \phi_2$. To ensure the two coordinates are different, Hall introduced the restriction that x be orthogonal to y. Again by Theorem 8.5, we see that the solution is then given by setting $x = \phi_2$ and $y = \phi_3$ or by taking a rotation of this.

According to Koren [46], one advantage of this approach is its ability to compute optimal layouts (according to specific requirements) and a short computation time as compared to other existing graph drawing algorithms. As in

Hall, Koren also based his drawing by solving some constrained minimization problem of an energy function. The solution of the minimization determines the optimal drawing layout. He also introduced the following constrained energy minimization problem for weighted graphs:

$$\begin{aligned} \min_{x^1,\ldots,x^p\in\mathbb{R}^n} \quad & E(x^1,\ldots,x^p) = \frac{\sum_{i\sim j} w_{ij} d_{ij}^2}{\sum_{i<j} d_{ij}^2}, \\ \text{subject to:} \quad & \mathrm{Var}(x^1) = \mathrm{Var}(x^2) = \cdots = \mathrm{Var}(x^p), \\ & \mathrm{Cov}(x^i, x^j) = 0, \quad 1 \le i \ne j \le p, \end{aligned} \tag{8.13}$$

where $\mathrm{Var}(x)$ is the variance of x, defined by

$$\mathrm{Var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x(i) - \bar{x})^2, \tag{8.14}$$

and $\bar{x}$ is the mean of x,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x(i). \tag{8.15}$$

$Cov(x^k, x^l)$ is the covariance of x^k and x^l defined as

$$Cov(x^k, x^l) = \frac{1}{n}\sum_{i=1}^{n}(x^k(i) - \bar{x}^k)(x^l(i) - \bar{x}^l), \tag{8.16}$$

and d_{ij} is the Euclidean distance between nodes i and j in the p-dimensional layout

$$d_{ij} = \sqrt{\sum_{i=1}^{p}(x^k(i) - x^k(j))^2}\,. \tag{8.17}$$

Also it can be shown that for any $x \in \mathbb{R}^n$,

$$\sum_{i\sim j} w_{ij} d_{ij}^2 = \sum_{k=1}^{p} (x^k)^T L x^k.$$

To see this, first note that since W is symmetric and $D_{ii} = \sum_j w_{ij}$, we have

$$\begin{aligned} \sum_{i,j}(x(i) - x(j))^2 w_{ij} &= \sum_{i,j}(x^2(i) + x^2(j) - 2x(i)x(j))w_{ij} \\ &= \sum_i x^2(i) D_{ii} + \sum_j x^2(j) D_{jj} - 2\sum_{i,j} x(i)x(j) D_{ij} \\ &= 2x^T L x. \end{aligned}$$

As we mentioned earlier, the energy and the constraints are invariant under translation. This enables us to eliminate the degree of freedom by requiring that, for each $1 \leq k \leq p$, the mean of $x^k = 0$, i.e.,

$$0 = \sum_{i=1}^{n} x^k(i) = (x^k)^T.\mathbf{1}_n.$$

This requirement further simplifies the above constraints. For instance, the constraint $\mathrm{Cov}(x^k, x^l) = 0$ is equivalent to requiring the vectors to be pairwise orthogonal $(x^k)^T x^l = 0$, and the variance of the vector x^k is reduced to $\mathrm{Var}(x^k) = \frac{1}{n}(x^k)^T x^k$, since for each k the mean of $x^k = 0$ implies

$$\begin{aligned} 0 = Cov(x^k, x^l) &= \frac{1}{n}\sum_{i=1}^{n}(x^k(i) - \bar{x}^k)(x^l(i) - \bar{x}^l) \\ &= \frac{1}{n}\sum_{i=1}^{n} x^k(i)x^l(i) = (x^k)^T x^l,\ k \neq l \end{aligned}$$

and

$$\begin{aligned} Var(x^k) &= \frac{1}{n}\sum_{i=1}^{n}(x^k(i) - \bar{x}^k)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(x^k(i))^2 \\ &= \frac{1}{n}(x^k)^T x^k. \end{aligned}$$

Now, the uniform variance constraint can be simplified as
$\|x^1\| = \cdots = \|x^p\|$.

Remark: Requiring the drawing to have a finite variance, $Var(x) = c > 0$, where c is a constant, guarantees that the nodes are well scattered. In addition, the uniform variance constraint forces the nodes to be equally scattered along each of the axes, which further makes the drawing have a balanced aspect ratio. However, the choice of the scalar c is arbitrary and merely determines the length of x and the scale of the energy E. Therefore, we can take $c = 1$, without loss of generality. The second constraint (i.e., $Cov(x^i, x^j) = 0,\ 1 \leq i \neq j \leq p$),

ensures that there is no correlation between the axes, so that each additional dimension will provide us with as much new information as possible.

Combining all these, the last minimization problem can be written in the following simple (but equivalent) form,

$$\begin{aligned} \min_{x_1,\ldots,x_p\in\mathbb{R}^n} \quad & \frac{\sum_{k=1}^{p}(x^k)^T L x^k}{\sum_{k=1}^{p}(x^k)^T x^k}, \\ \text{subject to:} \quad & (x^k)^T x^l = \delta_{kl}, \quad k,l = 1,\ldots,p, \\ & (x^k)^T.1_n = 0, \quad k = 1,\ldots,p. \end{aligned} \tag{8.18}$$

As we discussed before, the theorems mentioned earlier in this section are valid for any symmetric semi-positive definite matrix A and any positive definite matrix B. Now substituting $B = I_n$, the $n \times n$ identity matrix, in Theorem 8.8, and using the fact that the lowest eigenvector of the Laplacian L is $\mathbf{1}_n$, a p-dimensional optimal layout of a graph G is given by the lowest positive Laplacian eigenvectors $\phi_2, \ldots, \phi_{p+1}$ of the Laplacian matrix of G. Therefore, a p-dimensional layout of a graph uses eigenvectors $\phi_2, \ldots, \phi_{p+1}$ of the Laplacian matrix of G as coordinates of the graph. The coordinates of node i are given by $(\phi_2(i), \ldots, \phi_{p+1}(i))$. For example, a 2-dimensional drawing is obtained by taking the x-coordinates of the nodes to be given by ϕ_2, and the y-coordinates to be given by ϕ_3. Further, notice that as the eigenvector of eigenvalue zero is the constant $\mathbf{1}_n$, it will not be very useful for drawing the graph. Now let us see some examples of graph drawing demonstrating the eigenprojection method.

Example 1. One of the simplest graphs is the path graph. In figure 8.1, we plot the path graph on 20 nodes using the second eigenvector of the graph Laplacian. The x-axis is the number of the vertex (which corresponds to each vertex $v \in V(G)$) and the y-axis is the value of the eigenvector at that vertex.

Example 2. Consider a cycle graph on 300 vertices. Since we cannot draw a cycle graph in one dimension, we will need at least two eigenvectors. We plot the graph in two dimensions by using the second and third eigenvectors of the

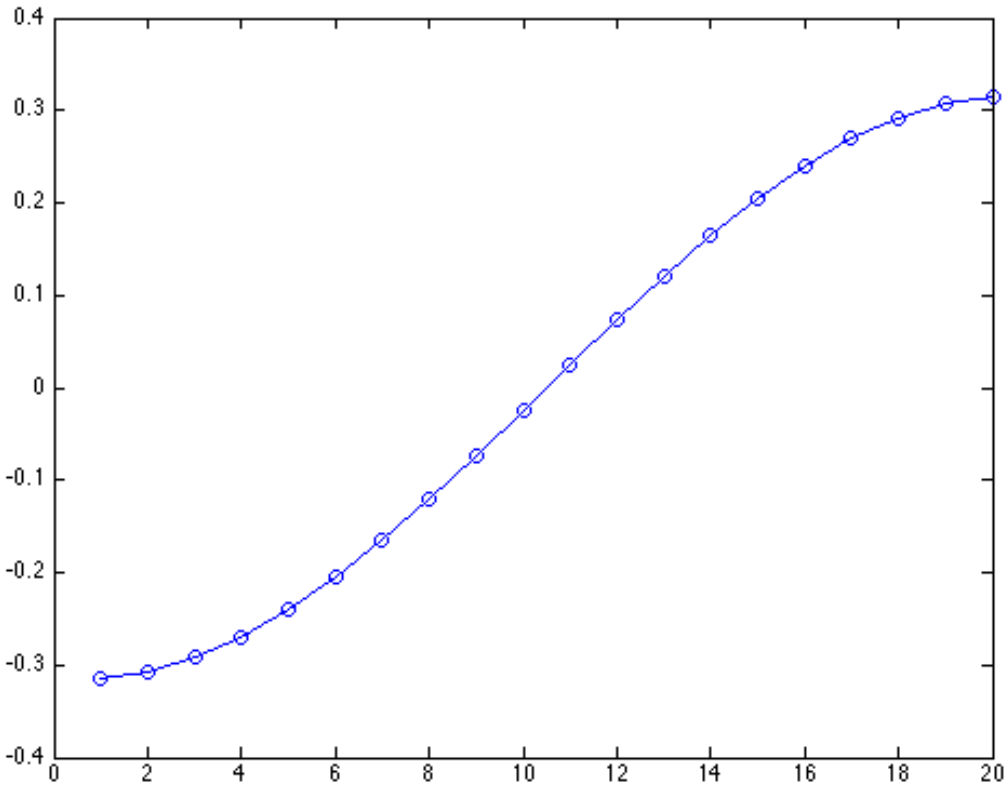


Figure 8.1: Drawing a path graph using the 2nd eigenvector of the Laplacian

Laplacian. In this case we draw each vertex i at coordinate $(v_2(i), v_3(i))$ (fig. 8.2) and $(v_2(i), v_5(i))$ (fig 8.3). Notice that in figure 8.3, we used the 2nd and 5th eigenfunctions to plot the same cycle graph and the outcome is not really a cycle graph.

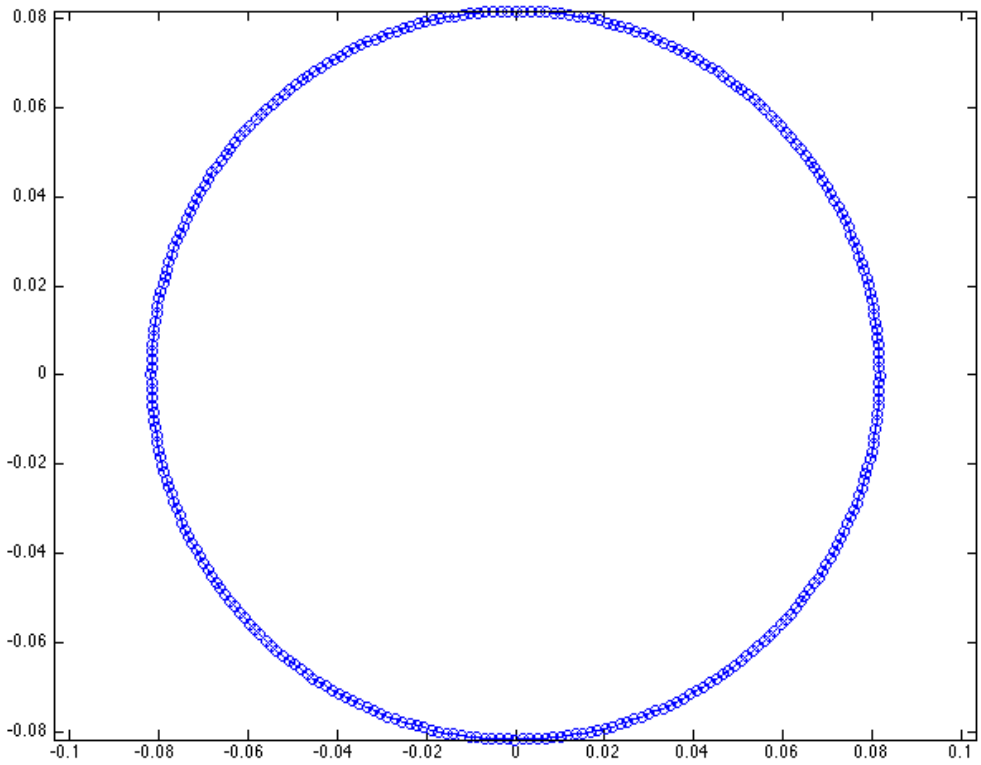


Figure 8.2: Drawing a cycle graph using the the 2nd and 3rd eigenvectors of the Laplacian

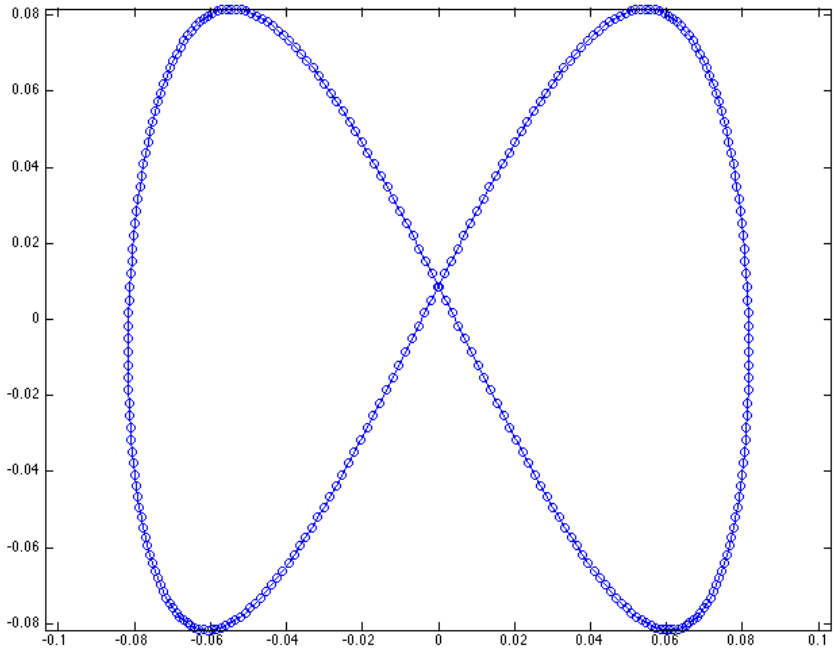


Figure 8.3: Drawing a cycle graph using the 2nd and 5th eigenvectors of the Laplacian

Example 3. In this example we generated the Delaunay graph of 500 randomly chosen points in the unit square (left) and we then plotted the same graph by using the 2nd and 3rd eigenvectors of the Laplacian (right).

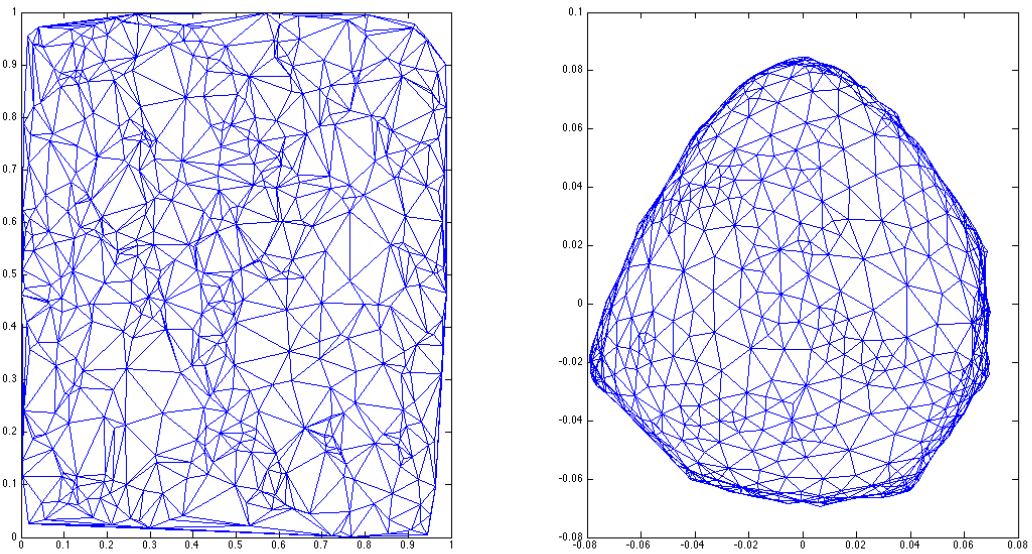


Figure 8.4: Drawing the Delaunay graph using the eigenvector of the Laplacian

It is also possible to use the largest eigenvectors (eigenvectors corresponding to the largest eigenvalues) of the adjacency matrix instead of the Laplacian. For instance, Fowler and Manolopoulos [59] used the eigenvectors of the adjacency matrix to draw molecular graphs. In a recent work of Brandes and Willhalm [60], eigenvectors of a modified Laplacian were used for the visualization of bibliographic networks.

8.4 Drawing graphs using splines

In chapters 5 and 6, we have seen that eigenfunctions of the graph Laplacian corresponding to the smaller eigenvalues can be approximated by the so-called interpolating variational splines on a given subset (sampling set) of the graph. We also have seen the eigenprojection method, which proves that

those eigenfunctions can be used to layout graphs. So it seems plausible to expect that splines approximating these eigenfunctions could also be used to draw graphs. Of course, since eigenfunctions and the approximating splines coincide only on a small subset of the graph, one should not expect a drawing to be as good as the drawing obtained using eigenfunctions. However, one can achieve a reasonable drawing by increasing the density of the sampling set or by choosing an appropriate degree of the spline.

The motivation behind using splines to draw graphs is simple. First, we needed spline approximations of eigenfunctions mostly because eigenfunctions are not localized in the sense that it is difficult to find a proper subset U of $V(G)$ which is considerably smaller than $V(G)$ such that an eigenfunction vanishes on U or is "very small" on the complement of U. Second, computing eigenfunctions is either costly or difficult for large graphs. Third, the dimension of spline spaces is much more smaller than the function space $L_2(G)$ and hence can be used in dimension reduction.

Example 4. Consider a cycle graph on 100 vertices. We take 20% of the vertex set for the sampling set (bottom left) and only 10% (bottom right) and compute the splines interpolating the second and third Laplacian eigenfunctions. Then we plot the graph by using the splines instead of the eigenfunctions. We get a remarkably nice plot when the density of the sample set increases as shown in the following figures.

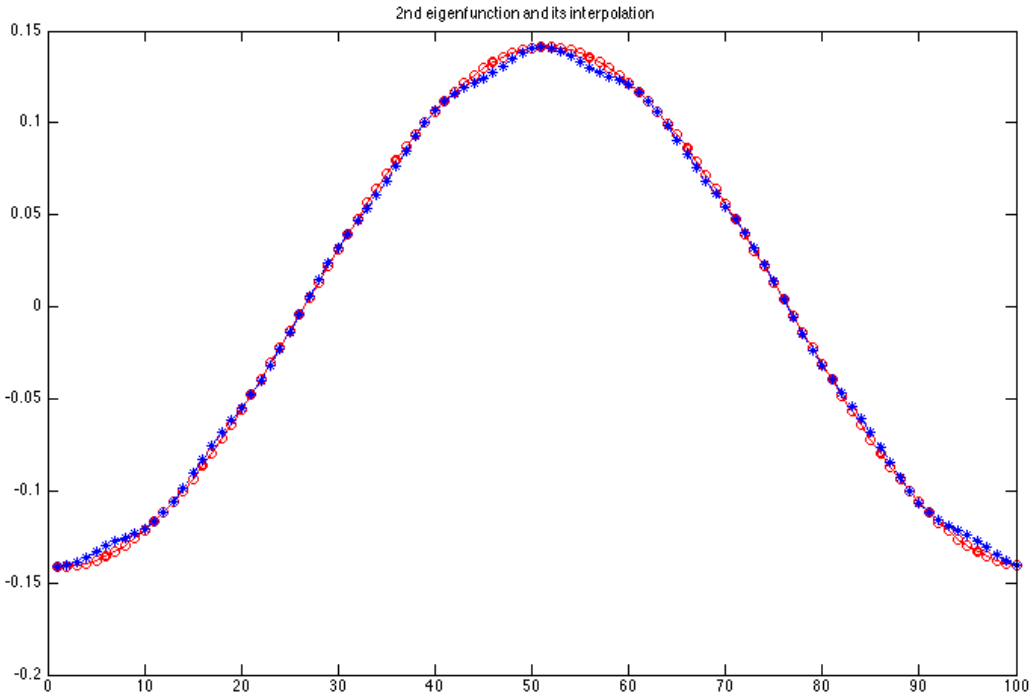


Figure 8.5: Approximating the 2rd eigenfunction (red) using splines (blue), sample 10%, k=15

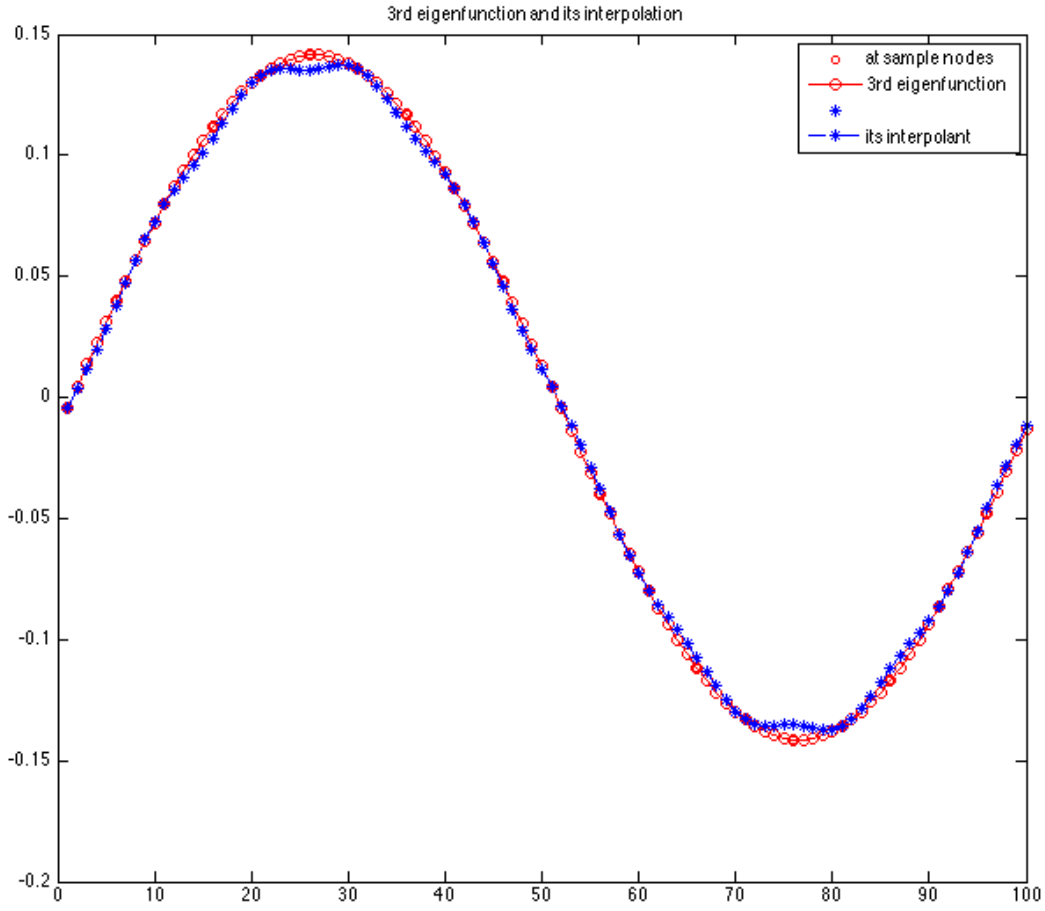


Figure 8.6: Approximating the 3rd eigenfunction (red) using splines (blue), sample 10%, k=15

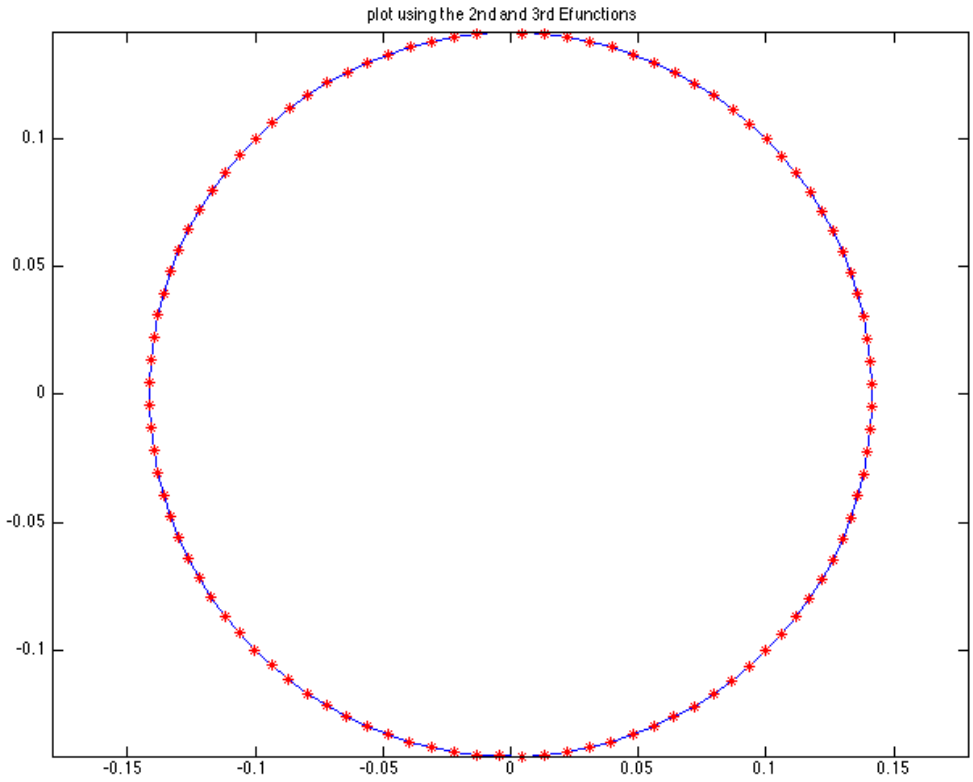


Figure 8.7: Drawing a cycle graph using 2nd and 3rd eigenfunctions

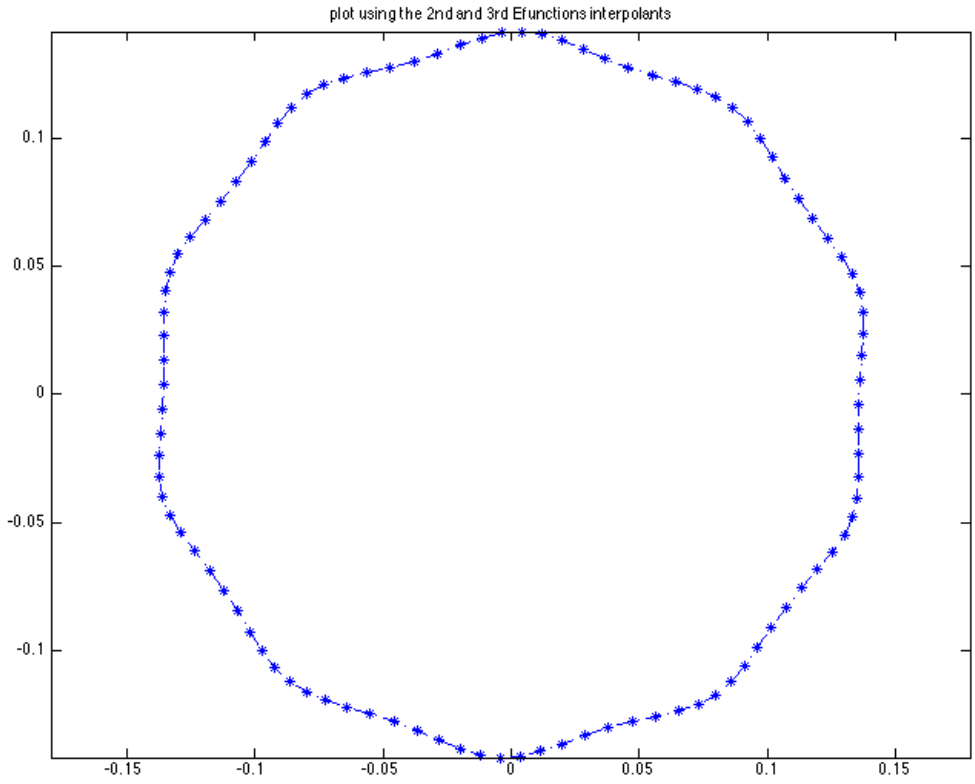


Figure 8.8: Drawing a cycle graph using splines

Example 5. We plot a 20 by 20 two dimensional grid. We took only 20% uniformly distributed vertices for the sampling set.

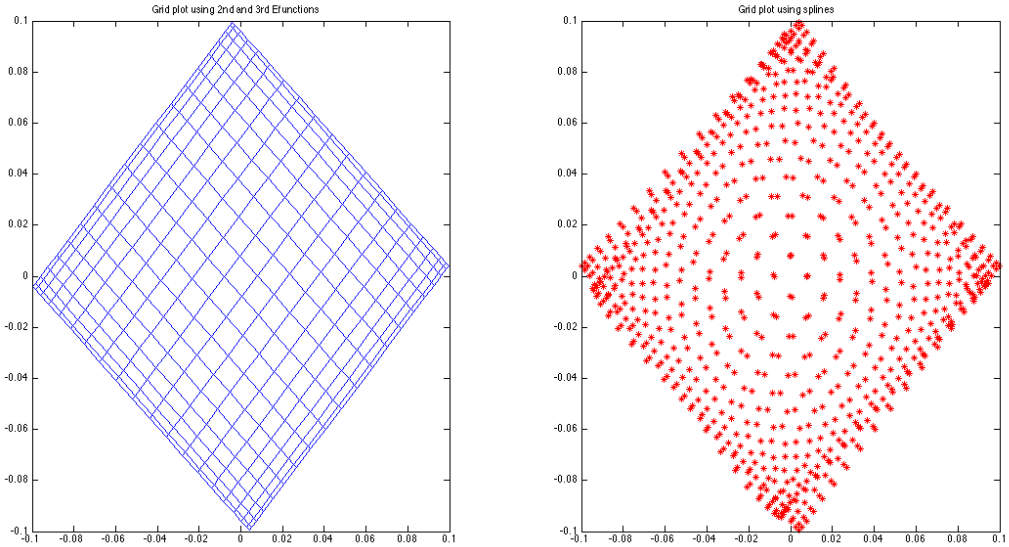


Figure 8.9: Drawing a 2D grid using splines, sample 20%, k=15.

We will end this section by reminding that even if eigenvectors (and splines) are the natural choices to draw graphs, it should come with the warning that they do not produce aesthetically nice pictures of all graphs. In fact, they produce bad pictures of most graphs. But, it is still the first thing we might try when we encounter a new graph that we want to understand.

8.5 Embedding graphs into high-dimensional space

Representing a graph on n vertices in Euclidean space $\mathbb{R}^m$, for $m < n$ has become a common strategy in graph drawing for visualization and other purposes, as discussed in the previous section. By setting the problem as minimizing Hall's energy function associated to a representation, a formulation of the problem for which the optimal solution can be simply computed in terms

of the lowest eigenvectors of the graph Laplacian (eigenprojection method). On the other hand, one can visualize and understand the structure of a graph if they are represented in low dimensions. Hence drawing graphs in low dimension is a convenient way for the same purpose. However, it is not always possible to achieve certain aesthetic properties in low dimensions. Aesthetics are attributes that define a "good" graph. Some of the qualities which determine a good graph are minimize edge crossing, minimize sum of edge lengths, minimize area, etc. Thus, embedding a graph in high dimensions gives us more flexibility and room to achieve the desired properties.

For example, in figure 8.10 (left) we draw a Buckyball in two dimensions by using the second and third eigenfunctions of the Laplacian, and we draw the same ball in three dimensions by using the second, third, and fourth eigenfunctions of the Laplacian (right). In the first picture we observe that the picture looks like a squashed Buckyball. The reason is that there is no canonical way to choose the eigenvectors v_2 and v_3. The smallest non-zero eigenvalue ($\lambda_2 = 0.2434$) of the Laplacian has multiplicity three. This graph should be really drawn in three dimensions, using any set of orthonormal vectors v_2, v_3, v_4 of the smallest non-zero eigenvalue of the Laplacian. In this case we obtain the standard embedding of the ball in $\mathbb{R}^3$.

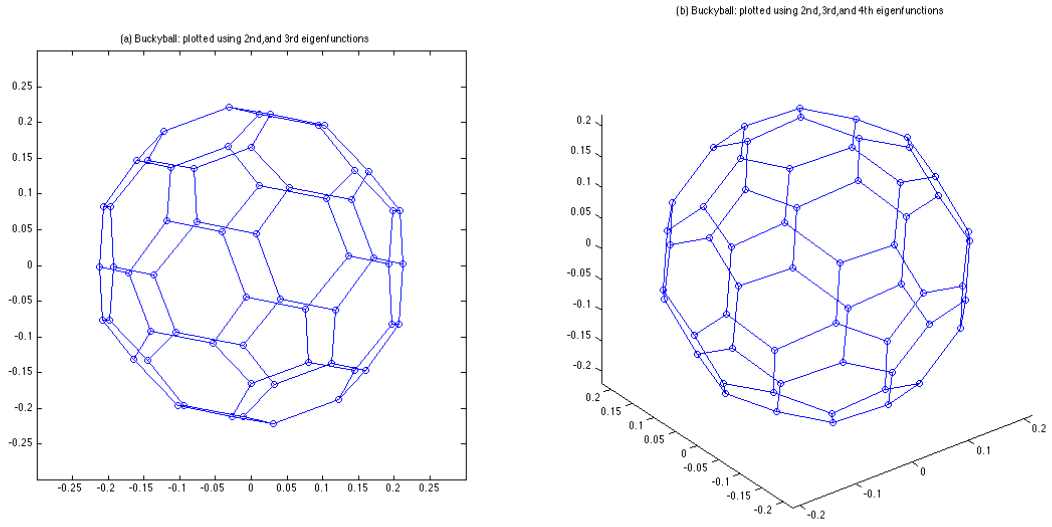


Figure 8.10: Buckyball in 2D (left) and 3D (right)

Another approach of graph drawing, which is similar to an m-dimensional layout of the graph is the High-Dimensional Embedding (HDE) [47]. The high-dimensional embedding method involves a two-step process: It first embeds a graph into high-dimensional space (e.g., in 50 dimensional space) and then projects the layout back to low-dimensional space (usually two or three) using a well-known multivariate analysis technique called *principal component analysis (PCA)*. In order to embed a graph in the m-dimensional space, we choose m *pivot nodes* that are almost uniformly distributed on the graph and associate each of the m axes with a unique node. The axis X^i, which is associated with pivot node p_i, represents the graph from the "viewpoint" of node p_i. This is done by assigning the j-th component of X^i to the graph-theoretic distance between nodes p_i and j (*i.e.*, $X^i(j) = d(p_i, j)$). Hence p_i is located at place 0 on axis X^i, its immediate neighbors are located at place 1 on this axis, and so on.

The graph theoretic distances are computed using breadth-first-search (BFS). Given an initial vertex $v \in V(G)$, the BFS algorithm traverses all the vertices

reachable from v and outputs those vertices in ascending order of their distances to v. The pivots $p_1, p_2, \ldots, p_m$ are chosen as follows. The first node, p_1, is chosen at random. For $j = 2, \ldots, m$, node p_j is a node that maximizes the shortest distance from $\{p_1, p_2, \ldots, p_{j-1}\}$. Now let the pivots be the set $\{p_1, p_2, \ldots, p_m\} \subset V(G)$. Each node $v \in V(G)$ is associated with m coordinates $(X^1(v), X^2(v), \ldots, X^m(v))$, such that $X^i(v) = d(p_i, v)$.

Once the graph is embedded in $\mathbb{R}^m$, we then project the coordinates into lower dimensions by using principal component analysis, which involves computing the first few largest eigenvalues and eigenvectors of the *covariance matrix* of points in the higher dimension. PCA transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components (PCs). The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. By using only the first few principal components, PCA makes it possible to reduce the number of significant dimensions of the data, while maintaining the maximum possible variance thereof. See [54] for a comprehensive discussion of PCA.

The next section presents the construction of principal components using the PCA technique to project a high dimensionally embedded graph into a lower dimension.

8.6 Projecting a high-dimensional embedding into low-dimension

The high-dimensional embedding method embeds a graph in high-dimensional space and then projects it back to low-dimensional space. Once we embed the graph into high dimension $\mathbb{R}^m$, the next step is to project the graph into lower dimension $\mathbb{R}^k$, $k < m$ (usually k=2 or 3). Now, assume that we have the m axes $X^1, \ldots, X^m$ (X^i are n-dimensional, $n = |V(G)|$), describing all the vertices of the graph. We want to represent the n vertices using only the k-

dimensions, by using k uncorrelated vectors say $Y^1, \ldots, Y^k$ (remember Y^i are also n-dimensional), which are the principal components. This time the coordinates of vertex v are $(Y^1(v), \ldots, Y^k(v))$. Each of the principal components among $Y^1, \ldots, Y^k$ is a linear combination of the axes $X^1, \ldots, X^m$. i.e.,

$$Y^i = \sum_{j=1}^{m} \alpha_j X^j \qquad 1 \leq j \leq k$$

for some scalar α_j's.

Let m_i denote the mean of the X^i-th axis. Thus,

$$m_i = \frac{1}{n} \mathbf{1}_n^T X^i = \frac{1}{n} \sum_{j=1}^{n} X^i(j). \tag{8.19}$$

The first stage of the PCA is to balance the m-dimensional representation, i.e., to center the data around 0. This is just a translation and does not affect the drawing at all. Next denote the vectors of the centered data by $\hat{X}^1, \ldots, \hat{X}^m$, defined as

$$\hat{X}^i = X^i - \frac{1}{n} \mathbf{1}_n^T X^i, \tag{8.20}$$

i.e.,

$$\hat{X}^i(j) = X^i(j) - m_i, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n.$$

We then construct a new balanced representation by the $m \times n$ matrix X, whose rows are the centered coordinates $\hat{X}^i$, $1 \leq i \leq m$:

$$X = \begin{pmatrix} \hat{X}^1(1) & \cdots & \hat{X}^1(n) \\ \vdots & \vdots & \vdots \\ \hat{X}^m(1) & \cdots & \hat{X}^m(n) \end{pmatrix}. \tag{8.21}$$

Notice that the graph is still embedded in the m-dimensional space but we just relocated the vertices so that the center of gravity is at the origin (i.e., the representation is balanced).

Now define the *covariance matrix* S, of dimension $m \times m$ as

$$S = \frac{1}{n} X X^T. \tag{8.22}$$

Next, compute the largest k eigenvectors $\mathfrak{u}^1, \ldots, \mathfrak{u}^k$ of S (these are the eigenvectors corresponding to the largest eigenvalues) and normalize them, so that these k vectors are orthonormal. Clearly S is symmetric as

$$S^T = (\frac{1}{n} X X^T)^T = \frac{1}{n}(X^T)^T X^T = \frac{1}{n} X X^T = S,$$

and hence all its eigenvalues are real.

The last step is to compute the new coordinate axes of the desired low-dimensional space. The first new axis Y^1 is the projection of the data in the direction of $\mathfrak{u}^1$, the next axis Y^2 is the projection in the direction of $\mathfrak{u}^2$, and so on. Hence the new coordinates are defined by:

$$Y^i = X^T \mathfrak{u}^i, \quad i = 1, \ldots, k. \tag{8.23}$$

Each node $v \in V(G)$ is now associated with k coordinates $(Y^1(v), \ldots, Y^k(v))$.

8.7 Spectral graph drawing within spline spaces

So far we have seen some applications of splines in signal approximation and drawing simple graphs. In this section, we will extend their application to dimension reduction.

At this point, it is clear that graphs can be represented in high dimensional spaces by using eigenvectors of some matrices related to the graph. In particular, the graph Laplacian operator has played a significant role in graph drawing as discussed in the previous sections. For example, the high-dimensional embedding technique uses the first few eigenvectors of the Laplacian of G as coordinates of the graph. Today, the applications of graphs are everywhere and almost in every discipline (e.g., psychology, social study, medical science, biology, chemistry, mathematics, engineering, computer science, etc) and the number of vertices of a graph is increasing drastically. For example, consider the social network graph Facebook: vertices are people and the edges represent Facebook friendship. This social graph has already millions of vertices and edges and still expanding. So now it becomes clear that it is not an easy

task to compute even the first few eigenvectors of the Laplacian L if the number of vertices of the graph is very large (e.g., if the graph has 10^6 vertices). As a result, researchers are trying to find a way to avoid this difficult and costly computation by restricting the drawing axes to lie within a "suitable" lower dimensional subspace of $\mathbb{R}^n$.

Obviously, constraining the drawing to lie within a subspace may not provide the desired outcome. In fact, it might result in arbitrarily bad layout unless the subspace is chosen very carefully. Strictly speaking, when we say a drawing is constrained in a subspace we really mean the drawing axes lie in that subspace. That means we restrict the optimization problem on the subspace in question in stead of the whole space $\mathbb{R}^n$, $n = V(G)$. A particular example is the work of Koren [49]. His work focuses on the construction of an appropriate subspace with relatively low dimensionality that captures aesthetically nice layouts of the graph. Here, we propose spline spaces as natural choices for suitable subspaces of $\mathbb{R}^n$. Our main contribution lies in providing a new low dimensional subspace of $\mathbb{R}^n$ for the construction of the new drawing (coordinate) axes. Let us recall the construction of spline spaces.

Let $G = (V, E)$, be a graph with $|V(G)| = n$. Given a subset of vertices $\mathcal{U} \subset V(G)$, a sequence of real numbers $\mathbf{r} = \{r_u\},\ u \in \mathcal{U},$ a positive real number $t > 0$, and a small positive $\epsilon > 0$, consider the following variational problem:
Find a function $s \in L_2(G)$ which satisfies

1. $s(u) = r_u,\ u \in U.$
2. s minimizes the functional $g \to ||(\epsilon I + L)^t g||$.

It is known that for any sequence of values $(r_1, r_2, \ldots, r_m) = \mathbf{r}$, and any $t > 0$ the minimization problem has a unique solution, where $m = |\mathcal{U}|$. This unique solution is called a **variational spline of order** t, and we denote it by $s_t(\mathbf{r})$. We denote the set of all splines for a fixed set $\mathcal{U} \in V(G)$, and fixed numbers $t > 0, \epsilon > 0$ by $\mathfrak{S}(\mathcal{U}, t, \epsilon)$.

$\mathfrak{S}(\mathcal{U}, t, \epsilon)$ is a closed subspace of $L_2(G)$ and dim $\mathfrak{S}(\mathcal{U}, t, \epsilon) =$ —U—. In fact, if $\mathfrak{h}^\nu$ is a solution of the variational problem such that $\mathfrak{h}^\nu(u) = \delta_{\nu,u}$, $\nu, u \in \mathcal{U}$, where $\delta_{\nu,u}$ is the Kronecker delta , then $\mathfrak{S}(\mathcal{U}, t, \epsilon) = \text{span}\{\mathfrak{h}^\nu\}_{\nu\in\mathcal{U}}$. More precisely,

$$s_t(\mathbf{r}) = \sum_{\nu\in U} r_\nu \mathfrak{h}^\nu, \qquad r = \{r_u\}_{u\in\mathcal{U}}.$$

Suppose the vertices in $\mathcal{U}$ are chosen in such a way that they are uniformly distributed over the graph. Let $\mathcal{X}$ be the $n \times m$ matrix whose columns are $\mathfrak{h}^\nu, \nu \in \mathcal{U}$, where m is the number of vertices in $\mathcal{U}$. Then we have $\mathfrak{S}(\mathcal{U}, t, \epsilon) = \mathit{Range}(\mathcal{X})$. Such a matrix representation is very convenient, since it allows us to describe the vectors in the subspace as the matrix-vector product $\mathcal{X}y$, where $y \in \mathbb{R}^m$. It is also easy to see that $\mathcal{X}$ is orthogonal, i.e., $\mathcal{X}^T\mathcal{X} = I$.

We recall that in the eigenprojection method of graph drawing, the drawing axes were chosen arbitrarily in $\mathbb{R}^n$. But this time we would like to constrain the drawing axes to lie within $\mathfrak{S}(\mathcal{U}, t, \epsilon)$. Thus, we require $X^1, \ldots, X^p \in \mathfrak{S}(\mathcal{U}, t, \epsilon)$. Consequently, we can always denote the axes of the drawing by the vectors $y^1, \ldots, y^p \in \mathbb{R}^m$, so that

$$X^1 = \mathcal{X}y^1, \ldots, X^p = \mathcal{X}y^p.$$

Moreover, the eigenprojection method defines the axes as the minimizers of (8.18),

$$\begin{aligned} \min_{x^1,\ldots,x^p\in\mathbb{R}^n} \quad & \frac{\sum_{k=1}^p (x^k)^T L x^k}{\sum_{k=1}^p (x^k)^T x^k}, \\ \text{subject to:} \quad & (x^k)^T x^l = \delta_{kl}, \quad k, l = 1, \ldots, p, \\ & (x^k)^T . \mathbf{1}_n = 0, \quad k = 1, \ldots, p. \end{aligned}$$

So now we want to optimize $x^1, \ldots, x^p$ within the subspace $\mathfrak{S}(\mathcal{U}, t, \epsilon)$, instead of $\mathbb{R}^n$. This can be achieved by replacing them with $\mathcal{X}y^1, \ldots, \mathcal{X}y^p$. Hence (8.18) becomes,

$$\begin{aligned} \min_{y^1,\ldots,y^p\in\mathbb{R}^m} \quad & \frac{\sum_{k=1}^p (y^k)^T \mathcal{X}^T L \mathcal{X} y^k}{\sum_{k=1}^p (y^k)^T \mathcal{X}^T \mathcal{X} Y^k}, \\ \text{subject to:} \quad & (y^k)^T \mathcal{X}^T \mathcal{X} y^l = \delta_{kl}, \quad k, l = 1, \ldots, p. \end{aligned} \tag{8.24}$$

Remark: Since $\mathbf{1}_n \notin \mathit{Range}(\mathcal{X})$, the axes are not required to be orthogonal to $\mathbf{1}_n$ this time. Also, since $\mathcal{X}^T\mathcal{X} = I$, (8.24) can be further simplified as

$$\begin{aligned} \min_{y^1,\ldots,y^p \in \mathbb{R}^m} \quad & \frac{\sum_{k=1}^{p} (y^k)^T \mathcal{X}^T L \mathcal{X} y^k}{\sum_{k=1}^{p} (y^k)^T y^k}, \\ \text{subject to:} \quad & (y^k)^T y^l = \delta_{kl}, \quad k, l = 1, \ldots, p. \end{aligned} \tag{8.25}$$

Since the columns of $\mathcal{X}$ are linearly independent, the matrix $\mathcal{X}^T L \mathcal{X}$ is positive-definite. So by Theorem 8.8, the minimizers of (8.25) are the lowest eigenvectors of $\mathcal{X}^T L \mathcal{X}$. Therefore, the drawing can be achieved by first computing the eigenvectors corresponding to the lowest p eigenvalues of $\mathcal{X}^T L \mathcal{X}$, denoted by $v^1, \ldots, v^p$, and then taking the coordinate axes to be

$$X^1 = \mathcal{X} v^1, \ldots, X^p = \mathcal{X} v^p \in \mathfrak{S}(\mathcal{U}, t, \epsilon).$$

Conclusion: We extended the application of splines in spectral graph drawing. More specifically, we used splines approximating the lowest few eigenfunctions of the Laplacian to draw graphs and an explanation of why these splines may be used for graph drawing was given. The high-dimensional embedding, low-dimensional projection, and eigenprojection methods of graph drawing were discussed in detail. The high-dimensional embedding method involves a two-step process: It first embeds a graph into high-dimensional space (e.g., in 50 dimensional space) and then projects the layout back to low-dimensional space (usually two or three) using a well-known multivariate analysis technique called *principal component analysis (PCA)*. We then constructed a new low-dimensional subspace of $\mathbb{R}^n$ for the coordinate axes to project a graphs into low dimensions.

REFERENCES

[1] D. K. Hammond, P. Vandergheynst, and R. Gribonval, *Wavelets on graphs via spectral graph theory*, Appl. Comput. Harmon. Anal., vol. 30, no. 2, pp. 129150, Mar. 2011.

[2] R. R. Coifman and M. Maggioni, *Diffusion wavelets*, Appl. Comput. Harmon. Anal., vol. 21, no. 1, pp. 5394, 2006.

[3] J. C. Bremer, R. R. Coifman, M. Maggioni, and A. D. Szlam, *Diffusion wavelet packets*, Appl. Comput. Harmon. Anal., vol. 21, no. 1, pp. 95 112, 2006.

[4] A. D. Szlam, M. Maggioni, R. R. Coifman, and J. C. Bremer, Jr., *Diffusion-driven multiscale analysis on manifolds and graphs: topdown and bottom-up constructions*, in Proc. SPIE Wavelets, vol. 5914, Aug. 2005, pp. 445455.

[5] M. Gavish, B. Nadler, and R. R. Coifman, *Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning*, in Proc. Int. Conf. Mach. Learn., Haifa, Israel, Jun. 2010, pp. 367374.

[6] I. Daubechies : *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics Philadelphia, PA, 1992 .

[7] D. I. Shuman, S.K.Narang, P.Frossard, A.Ortega, and P.Vandergheynst: *The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains.* IEEE Signal Process. Mag., 30(3):83-98, 2013.

[8] M. Baker and R. Rumely : *Harmonic analysis on Metrized Graphs*, Canada J. Math. Vol.59 (2), 2007.

[9] F. Chung: *CBMS Lecture Notes on Spectral Graph Theory*, Amer. Math. Soc, Providence, RI, 1997.

[10] F. Chung and R. Richardson: *Weighted graph Laplacians and isoperimetric inequalities*, Pacific Journal of Mathematics 192 (2000), 257–273.

[11] F. Chung: *Laplacians and the Cheeger inequality for directed graphs*, Annals of Combinatorics, 9 (2005), 1–19.

[12] F. Bauer *Spectral Properties of the normalized graph Laplace operator of directed graphs.*

[13] A. Grigorian: *Analysis on graphs, Lecture notes*, University of Bielefed, WS 2011/12.

[14] C. D. Godsil and G. Royle: *Algebraic Graph Theory*, Springer-Verlag, New York, 2001.

[15] B. Mohar, W. Woess : *A survey on spectra of infinite graphs*, Bull. London Math. Soc. 21(1989), 209-234.

[16] B. Mohar: *Some Applications of Laplacian Eigenvalues of Graphs*:, G. Hahn, G. Sabidussi (Eds.), Graph Symmetry, Kluwer Academic Publishers, Dordrecht (1997), pp. 225275.

[17] B. Mohar: *The Laplacian spectrum of graphs, in: Graph Theory, Combinatorics, and Applications*, Wiley, 1991, pp. 871-898.

[18] B. Mohar: *The spectrum of an infinite graph*, Linear Algebra Appl. 48(1982) 245-256.

[19] A. Agaskar, and Y. M. Lu: *Uncertainty principle for signals defined on graphs: Bounds and characterizations,* in Proc. IEEE Int. conf., Speech, and Signal Process., Kyoto, Japan, 2012.

[20] D. Zhou, J. Huang, and B. Scholkopf: *Learning with hyper graphs: clustering, classification, and embedding*. In advance Neural Information Processing Systems (NIPS). MIT Press, Cambridge, MA, 2006.

[21] D. Zhou, J. Huang, and B. Scholkopf: *Learning from labeled and unlabeled data on a directed graph*. In Proc. 22th International Conference on Machine Learning, 2005.

[22] I. Pesenson: *Sampling in Paley-Wiener spaces on combinatorial graphs*, Trans. Amer. Math. Soc. 360 (2008), no. 10, 56035627.

[23] I. Pesenson: *Variational Splines and Paley-Wiener spaces on combinatorial graphs*, Constr. Approximation, 29 (2009), no. 1, 120.

[24] I. Pesenson: *Variational splines on Riemannian Manifolds with applications to Integral Geometry*, Apr 2011.

[25] I. Pesenson: *An approach to Spectral Problems on Riemanian Manifolds*, Pacific Journal of mathematics, May 2004.

[26] I. Pesenson: *Sampling of Band limited vectors*, J. of Fourier Analysis and application, 2001.

[27] I. Pesenson: *Removable sets and eigenvalue and eigenfunction approximations on finite combinatorial graphs, Applied and computational Harmonic Analysis*, Volume 29, Issue 2, Sept. 2010.

[28] I. Pesenson, M. Pesenson:*Eigenmaps and minimal and bandlimited immersions of graph into Euclidean Spaces*, Journal of Mathematical Analysis and Applications, Vol. 366, Issue 1, June 2010.

[29] I. Pesenson, H. Feichitinger:*A reconstruction method for bandlimited signals on the hyperbolic plane*, Sampling Theory, Signal and Image Processing 4(2005).

[30] I. Pesenson, H. Feichitinger:*Recovery of bandlimited signals on manifolds by an iterative algorithm, wavelets, frames, and operator theory*, 137- 152, Contemp. Math., 345, Amer. Math. Soc., Providence, RI, 2004.

[31] I. Pesenson, M. Pesenson, *Sampling, filtering and sparse approximations on combinatorial graphs*, J. Fourier Anal. Appl. 16 (2010), no. 6, 921-942.

[32] I. Pesenson, F. Hartmut: *Poincare and Plancherel-Polya Inequalities in Harmonic Analysis on Weighted Combinatorial Graphs*, SIAM J. Discrete Math. 27 (2013), no. 4, 2007-2028.

[33] A. I. Zayed: *On the notion of Bandlimittedness and its generalizations*, Rev. Union Math. Argent., 49(1):99-109, 2008.

[34] A. E. Brouwer and W.H. Hamers: *Spectra of Graphs*. Springer, Heidelberg, 2012.

[35] A. M. Duval, C. J.'Klivans, J. L. Martin: *Simplicial Matrix-Tree Theorems*,Trans. American Mathematical Society 361 (2009), no. 11, 60736114.

[36] J. M. Aldous, R. J. Wilson: *Graphs and Applications, An Introductory Approach*, Springer, 2000.

[37] T. Biyikoglu, J. Leydold and P. Stadler: *Laplacian Eigenvectors of Graphs; Perron-Frobenius and Faber-Krahn Type Theorems* Springer 2007.

[38] R. Varga, *Functional Analysis and Approximation Theory in Numerical Analysis*, SIAM, Philadelphia PA 1971.

[39] J. W. Jerome and L. L. Schumaker: *On Lg-Splines, J. Approximation Theory* (1969).

[40] R. M. Rustamov: *Average interpolating wavelets on point clouds and graphs*, Oct 2011.

[41] Ulrike Von Luxenberg: *A tutorial on Spectral Clustering*, statistics and computing, vol 17, 2007.

[42] A. K. Jain and R. C. Dubes: *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

[43] A. K. Jain, M. N. Murty, and P. J. Flynn: *Data clustering: a review*, ACM Computing Surveys, 31(3):264323, 1999.

[44] X. Zhu and M. Rabbat: *Approximating signals Supported on Graphs*, in proc. ICASSP, 2012.

[45] A. Sandryhaila and J. M. Moura: *Discrete Signal Processing on Graphs*, IEEE Trans. Signal Process., vol. 61, no. 7, pp. 1644-1656, 2013.

[46] Y. Koren: *Drawing Graphs by Eigenvectors; Theory and Practice*, Comput. Math. Appl. 49(2005).

[47] Y. Koren and D. Harel : *Graph Drawing by High-Dimensional Embedding*, Proc. 10th Graph Drawing(GD'02), LNCS 2528, pp. 207-219, Springer-Verlag, 2002.

[48] Y. Koren: *On Spectral Graph Drawing*, Proc. 9th Inter. Computing and Combinatorics Conference, LNCS, Springer -Varlag, 2003.

[49] Y. Koren: *Graph Drawing by Subspace optimization*, Proc. 6th Eurographics-IEEE TCVG Symposia on Visualization (VisSym'04), Eurographics, pp. 65-74, 2004.

[50] K. M. Hall: *An r-dimensional Quadratic Placement Algorithm*, Management Science 17(1970).

[51] G. Di. Battisa, P. Eades, R. Tamassia and G. Tollis: *Graph drawing Algorithms for the Visualization of Graphs*, Prentice Hall, 1999.

[52] M. Kaufman and D Wagner: *Drawing Graphs, Methods and Models*, Springer, 2001.

[53] H. Hosobe: *An Extended High-Dimensional Method for Interactive Graph Drawing*. In Proceedings of the Asia Pacific Symposium on Information Visualization Applied Computing (AVIS), 2004.

[54] B. S. Everitt and G. Dunn: *Applied Multivariate Data Analysis*, Arnold, 1991.

[55] T. H. Cormen, C. E. Leiserson and R. L. Rivest : *Introduction to Algorithms*, MIT Press, (1990).

[56] D. Spielman: *Spectral graph theory and its application*, Lecture Notes, Yale University, 2009.

[57] A. Civril, M. M. Ismail, E. B. Rivele: *Graph drawing using spectral distance embedding*. In GD'05, Limerick, Ireland (2005).

[58] Kobourov, G. Stephen: *Spring Embedders and Force-Directed Graph Drawing Algorithms*, CoRR **abs/1201.3011** (2012).

[59] P. Fowler and D. Manolopoulos: *Molecular Graphs, point groups and Fullerence*, J. Chem. Phys.96(1992), 7603-7614.

[60] U. Brandes and T. Willhalm: *Visualizing Bibliographic Networks with a reshaped Landscape Metaphor* Annals. of Combinatorics 9 (2005), 1-19.

[61] S. Butler: *Eigenvalues and Structure of Graphs, PhD. Dissertation, University of California, San Diago, 2008.*

[62] R. A. Brualdi and D. Cvetkovic: *A combinatorial Approach to Matrix Theory and its applications*, CRC press 2009 .

[63] P. V. Mieghem: *Graph Spectra for Complex Networks*, Cambridge 2011.

[64] A. N. Langville and C. D. Mayer: *Deeper Inside PageRank*, Oct 2004.

[65] O. Christensen: *An introduction to Frames and Riesz Bases*, 2008.

[66] K. Grochenig: *Foundations of Time-Frequency Analysis*. Appl. Numer. Harmon. Anal. Birkhauser Boston, 2001.

[67] K. Grochenig: *Localization of frame, Banach spaces, and the invertibility of the frame operator*. J. Four. Anal. Appl, 10(2), 2004.

[68] R. B. Bapat, J. W. Grossman and D. M. Kulkarani: *Generalized Matrix tree theorem for mixed graphs*, Linear and Multilinear Algebra **46**, pp.299-312.

[69] R. B. Bapat, S. Pati, and D. Kalita: *On weighted directed graphs*, Linear Algebra Appl,438 (2012), 99 -111.

[70] X. D. Zhang, J. S. Li: *The Laplacian spectrum of a mixed graph*, Linear Algebra and its Applications, 353(2002), 11-20.

[71] X. D. Zhang, R. Luo: *The Laplacian eigenvalues of a mixed graph*, Linear Algebra and its Applications, 362(2003), 109-119.

[72] R. Horn and C. Johnson: *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1999.

[73] F. Astengo: *An uncertainty principle on homogeneous Trees*, Proceedings of the Amer. math Soc., Vol131, no. 10, 2003.

[74] M. Ghoniem, A. Elmoataz, and O. Lezoray: *Discrete Infinity Harmonic Functions: Towards a Unified Interpolation Framework on Graphs*

[75] M. Keller: *The Essential spectrum of the Laplacian on rapidly branching tessellations*, Math. Ann.346 (2010), no. 1, 51-66.

[76] M. Keller, S. Haeseler, D. Lenz, and R. Wojciechowski: *Laplacians on Infinite Graphs; Dirichlet and Neumann Boundary Conditions*, JSpectr. Theory **2** (2012) 397-432.

[77] P. E. T. Jorgensen: *Essential Self-adjointness of the graph Laplacian*, Journal of Mathematical Physics, 49(2008), no, 7,073510.

[78] Y. Colin de Vediere, N. Torki-Hamza: *Essential self-adjointness for combinatorial Schrödinger operators II-Metrically non complete graphs*, Math. Phys. Anal. Geom. 14(2011), no. 1, 21-38.

[79] A. Weber: *Analysis of the physical Laplacian and the heat flow on locally finite graphs*, J. Math. Anal. Appl. 370(2010), no, 1, 146 -158.

[80] M. H. Quang, S. H. Kang, and T. M. Le: *Image and video Colorization using vector valued Reproducing Kernel Hilbert Spaces*, J.Math. Imaging Vision 37 (2010), pp. 49-65.

[81] N. Houhou, X. Bresson, et al: *Semi-supervised segmentation based on Non-local Continuous Min-Cut*,Proc. of the intl. Conf. on Scale Space and Variational Methods in Computer Vision, LNSC 5567, Voss, Norway, June 2009.

[82] C. E. Shannon : *Communication in the Presence of Noise*, in Proc. IRE, vlo.37 1949, pp. 10-21.

[83] M. Unser: *Sampling – 50 Years After Shannon*, in Proc. IEEE, vlo.88 No. 4, April 2000, pp. 569-587.

[84] Y. Liu: *Irregular Sampling for spline wavelets subspaces*, IEEE Trans. Inform. Theory, vol. 42, no. 2, pp. 623-627, 1996.

[85] Y. Lyubarskii and W. R. Madych: *The recovery of irregularly sampled bandlimited functions via tempered splines*, J. Functional Analysis, vol. 125, no. 1, pp. 201-222, 1994.

[86] G. B. Folland: *Fourier Analysis and its Applications*, Wadsworth & Brooks/Cole, Pacific Grove, Calif., 1992.

[87] S. K. Narang, A. Gadde, E. Sanou and A. Ortega : *Localized Iterative Methods for Interpolation in Graph structured Data*, **arXiv**:1310.2646.

[88] K. Gröchenig: *Reconstruction Algorithms in Irregular sampling*, Mathematics of computation, vol 59.199, July 1992.

[89] K. Gröchenig: *Foundations of time-frequency analysis*, Birkhäuser, Boston Inc., Boston, MA, 2001.

[90] S. K. Narang, A. Gadde and A. Ortega: *Signal Processing techniques for interpolation in Graph structured data*, ICASSP, May 2013.

[91] A. Sharma: *Application of Graph Theory in Operational Research Problems: Basic Concepts, formulation of the problems, Algorithms*, May 2012.

[92] G. Finke: Operational Research and Networks, Wiley, 2008.

[93] E. Gine and V. Koltchinskii: *Empirical graph Laplacian aproximation of Laplace-Beltrami operators*, arXive:math/0612777v1.

[94] S. Hoory, N. Linial, and A. Wigderson: *Expander Graphs and Their Applications*, AMS, vol 43.4, Oct 2006.

[95] D. Auber: *A huge graph Visualization framework, In Graph Drawing Software*, pp.105-126, Springer-Verlag, 2003.

[96] I. J. Schoenberg: *On Interpolation by Spline functions and its minimal properties*, On Approximation Theory (Proc. Oberwolfach Conf. 4-

10 Aug, 1963) (P.L.Butzer and J.korevaar,eds), ISNM Vol.5, Birkhauser, Basel, 1964, pp.109-129.

[97] I. J. Schoenberg: *Spline Interpolation and higher derivatives*, Proc. nat.Acad. Sci.,51(1964), pp.24-28.

[98] I. J. Schoenberg: *On best approximations of Linear Operators*, Koninkle.Ned. Akad. Wetenschap. Proc. Ser. A, 67 (1964), pp. 155-163.

[99] J. J. Molitierno: *Applications of combinatorial matrix theory to Laplacian Matrices of Graphs*, Taylor & Francis (2012).

[100] R. J. Duffin and A. C. Schaeffer: *A class of non harmonic Fourier series*, Trans. Amer.Mathe. Soc., vol. 72, pp.341-366,1952.

[101] R. G. Wiley: *Recovery of bandlimited signals from unequally spaced samples*, IEEE Trans. Comm. 26, 135-137, (1978).

[102] P. L. Butzer and J. Lei: *Approximation of signals using measured sampled values and error analysis*, communication in Applied Analysis 4, 245-255, (2000).

[103] H. Feichtinger and K. Grochenig: *The theory and practice of irregular sampling*, in Wavelets: Mathematics and Applications, pp. 305-363, CRC Press, (1994).

[104] M. Gavish, B. Nadler, R. Coifman: *Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi-supervised learning*, ICML 2010.

[105] R. Wojciechowski: *Heat kernel and essential spectrum of infinite graphs*, Indiana Univ. Math. J. 58 (2009), no. 3, 1419-1441.

[106] T. Washio and H. Motoda:*state of the art of graph-based data mining.* ACM SIGKDD Explorations Newsletter, 5(1):59-68, 2003.4.

[107] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, and C. J. Stam: *The application of graph theoretical analysis to complex networks in brain*, Clinical Neurophysiology, vol.118, no.11, pp. 2317-2331, 2007.