University of Padova

Department of Mathematics "Tullio Levi-Civita"

Master Degree in Data Science

Interpolating Graph Signals using Positive Definite Graph Basis Functions

Supervisor
Erb Wolfgang
University of Padova

Master Candidate
Simone Bianchi

Academic Year

2023-2024

Acknowledgements

I would like to express my deepest gratitude to my thesis supervisor, Wolfgang Erb, for his invaluable guidance, constant support, and insightful feedback throughout the process of this work.

I am deeply grateful to my family for their unwavering love, understanding, and encouragement throughout the course of my studies. Their constant belief in my abilities, especially during moments of doubt and difficulty, has been an invaluable source of strength. Whether celebrating milestones or navigating challenges, their support has remained steadfast. Their faith in my potential has driven me forward, reminding me that no obstacle is insurmountable with their support. I am truly thankful for their sacrifices and for the enduring love that has carried me through this academic endeavor.

To my friends, thank you for being a constant source of support and laughter throughout this journey. Your encouragement during challenging times and your ability to bring joy and lightheartedness when I needed it most have been truly invaluable. Whether through a simple message, a shared laugh, or a moment of understanding, you have helped me stay grounded and reminded me of the importance of balance and connection. Your friendship has added so much meaning and richness to my life, making even the toughest days more manageable and the successes more fulfilling. I am forever grateful for your presence in my living and for the memories we have created together along the way.

Last but certainly not least, a special thanks to the Kart "Brotherhood" for the endless laughter and for always reminding me of the value of taking a break and having fun. Your support has been an unexpected yet deeply appreciated source of motivation throughout this journey

Abstract

To interpolate graph signals using generalized shifts of a graph basis function (**GBF**), we introduce the concept of **positive definite** functions on graphs. This concept blends kernel-based interpolation with **spectral theory** on graphs and can be seen as the graph equivalent of radial basis function interpolation in Euclidean spaces or spherical basis functions. We present several descriptions of positive definite functions on graphs, the most relevant one is a **Bochner type characterization** involving positive **Fourier coefficients**. These descriptions enable us to construct GBF's and delve deeper into GBF interpolation: we can characterize the **native spaces** of the interpolants, we provide explicit **estimates** for **interpolation error** and infer bounds for **numerical stability**. Finally, we show an application where GBF interpolation is used to obtain **quadrature formulas** on graphs.

Contents

1

Introduction

Graph signal processing is an expanding research area focused on analyzing large data structures within complex and irregular graph domains [1],[2],[3]. In today's world, such data structures are prevalent and amplified in various aspects of our lives, including social networks, healthcare systems, banking and shopping applications, as well as traffic and security monitoring. Graphs provide a means to model, connect, and organize these structures. In addition, graph signal processing provides tools to filter and streamline the data on these graphs, enabling the extraction of the most relevant information. For signals on complex graphs, interpolation and approximation techniques are crucial tools for minimizing computational costs or reconstructing signals from limited measurements. Inspired from classical signal processing on the real line, researchers have introduced spaces of bandlimited signals on graphs to approximate signals [4]. Similar to classical Fourier analysis, the core concept is that smooth signals can be accurately approximated using their low-frequency components, while the higher frequencies typically contain noise. As a result, domains of bandlimited signals serve as inherent approximation spaces on graphs. Nevertheless, employing bandlimited signals for graph signals interpolation presents some limitations:

- **(L1)** For a given set of interpolation nodes on the graph, the **uniqueness of interpolation** cannot be assured in the space of bandlimited functions, as referenced in [5].
- **(L2)** More crucial for applications, even when unisolvence is given, bandlimited interpolants exhibit unstable and highly oscillatory behavior, especially in the boundary regions of

the graph. Similar to the behavior observed in high-order polynomial interpolation, this can be viewed as a **Runge-type phenomenon**.

Similar to high-order polynomial interpolation, there are two methods to circumnavigate around the issues **(L1)** and **(L2)**:

(S1) **Adaptive selection of optimal sampling nodes**. This approach stabilizes the interpolation when the sampling nodes can be freely selected. For adaptive refinement, a typically expensive greedy algorithm is employed. Implementations of this strategy on graphs can be found in [6], [7], [8].

(S2) **Regularization of the interpolation conditions**. If the interpolation condition is relaxed to solving a regression problem, the target space of bandlimited signals can be shrunk, ensuring unisolvence **(L1)** [4], [7]. Alternatively, when the problem is formulated as a minimization problem with additional smoothness or sparsity constraints, as referenced in [3], [8], unisolvence is achieved, and the Runge artifacts described in **(L2)** are generally reduced.

Beyond the strategies **(S1)** and **(S2)**, kernel-based methods offer a more flexible approach for reconstructing signals from a limited set of samples. On graphs, these kernel methods are often explored through regression and regularization techniques for machine learning, as referenced in [4], [9], [10], [11], [12]. Typically used kernels include powers of the graph Laplacian [4] and diffusion kernels [10], [12]. Studies related to graph signal interpolation have emphasized variational splines as kernels, as seen in [13], [14].
For graph interpolation, a significant benefit of kernel-based methods is that unisolvence **(L1)** is automatically ensured when the applied kernel is positive definite. While specific kernels can be designed to emulate interpolation in bandlimited spaces, **(L2)** also influences kernel-based methods. However, the significant flexibility in generating multiple kernels allows for the customization of interpolation kernels to suit the given data, thereby preventing reconstruction artifacts. An example of a diffusion kernel based graph interpolation is shown in the figure (1.1). Choosing suitable interpolation kernels is crucial for the quality of signal reconstruction. Since the space consisting of all linear kernels grows quadratically with the number of graph nodes, It's important to concentrate on simpler classes of kernels. These should provide sufficient smoothness properties for interpolation and align well with the spectral structure of the graph. In Euclidean spaces, a class of kernels is represented by positive definite radial basis functions (RBF's) [15], [16], [17]. A radial basis function f naturally produces a symmetric kernel K in $\mathbb{R}^d$ by considering the shifts $K(x, y) = f(x - y)$ of the function f. This enables a compact

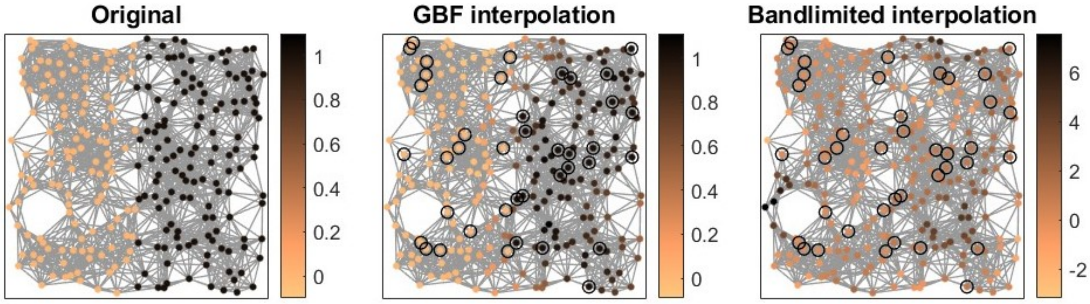


Figure 1.1: Comparison between GBF interpolation and bandlimited spectral interpolation on a sensor graph. Left: Original signal. Middle: Signal interpolated using GBF. Right: Signal interpolated using bandlimited interpolation, exhibiting strong Runge-type artifacts. The data samples are collected from the circled nodes of the graph.

representation of the kernel K in terms of a univariate function. Similar concepts of positive definite basis functions are found in other group settings, such as for periodic functions [18] or more generally on compact groups [19]. Furthermore, they are recognized in symmetric spaces such as the unit sphere. In this scenario, the associated positive definite functions that yield the kernels are known as spherical basis functions (SBF's) [20], [21]. Positive definite functions and their extensions have a deep and extensive mathematical history, playing fundamental roles in harmonic analysis, signal processing, and probability theory [22], [23]. The central characterization of positive definite functions is provided by Bochner's Theorem [24], which connects positive definiteness in the spatial domain to positivity in the Fourier domain. This connection is essential for several applications. In signal processing, the standard kernels for convolution and signal filtering are primarily based on positive definite functions, such as the sinc filter and the Gaussian filter. In addition, for RBF's and SBF's, the positive definiteness of the basis function guarantees the positive definiteness of the corresponding kernel K, thereby ensuring the unisolvence **(L1)** of the interpolation problem. In the growing field of graph signal processing, a comprehensive theory of positive definite functions has not yet been explored. Thus, the aim of this study, based on the work of the article *"Graph signal interpolation with Positive Definite Graph Basis Functions"* by Wolfgang Erb [25], is to introduce and support the concept of positive definite functions on graphs and to examine their role as generators of kernel-based interpolation methods. In analogy to RBF interpolation in $\mathbb{R}^d$, our purpose is to understand how interpolation using generalized shifts of a positive definite graph basis function (GBF) can be implemented on graphs. Moreover, leveraging the Fourier properties of the positive definite basis function, we aim to provide further insights about the approximation capabilities and the stability properties of the GBF interpolation scheme.

2

Theoretical Landscape On Graph

2.1 Introduction to spectral graph theory

We start our analysis by providing a broad introduction to spectral graph theory, delving into key concepts such as the graph Fourier transform, graph spectrum, and graph convolution. A significant reference for spectral graph theory is the monography [26] by F. Chung. For an overview of the graph Fourier transform, graph convolution, and space-frequency concepts related to graphs, we refer to [27].
In our analysis, the graph G is define as a triplet $G = (V, E, \mathbf{A})$, where:

- $V = \{v_1, \dots, v_n\}$ represent a finite set of vertices.
- $E \subseteq V \times V$: set of edges (directed or undirected).
- $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a weighted, symmetric and non negative adjacency matrix containing the connection weights of the edges.

The harmonic structure of graph G is captured within the adjacency matrix $\mathbf{A}$. The harmonic structure of G is intrinsically undirected, even if the edges of G are directed, since $\mathbf{A}$ is assumed to be symmetric.
The aim of this study is to investigate interpolation of signals $x : V \to \mathbb{R}$ on the graph G. Let's represent the vector space containing all signals on G as $\mathcal{L}(G)$. Since the number of nodes in G is fixed, the size of $\mathcal{L}(G)$ is finite and equal to the number n of nodes. Given that the node set

V is ordered, we can represent the signal x as a vector $x = (x(v_1), \ldots, x(v_n))^{\mathrm{T}} \in \mathbb{R}^n$. Based on the context, we will interchangeably use the representation of x as a function in $\mathcal{L}(G)$ or as a vector in $\mathbb{R}^n$. In $\mathcal{L}(G)$ space, there is an intrinsic inner product given by

$$y^{\mathrm{T}}x := \sum_{i=1}^{n} x(v_i)y(v_i).$$

The associated Euclidean norm is defined as $\|x\|^{\,2} := x^{\mathrm{T}}x = \sum_{i=1}^{n} x(v_i)^2$. The canonical orthonormal basis in $\mathcal{L}(G)$ is denoted by $\{e_1, \ldots, e_n\}$ and given by the unit vector e_j satisfying $e_j(v_i) = \delta_{ij}$, the delta's Kronecker, for $i, j \in \{1, \ldots, n\}$, that is

$$e_j(v_i) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

Now, let's define the graph Laplacian $\mathbf{L} \in \mathbb{R}^{n \times n}$, which is related to the adjacency matrix $\mathbf{A}$ in order to establish a spectral structure for G

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \;\; \mathbf{L}_{ij} = \begin{cases} \mathbf{D}_{ij}, & \text{if } i = j \\ -\mathbf{A}_{ij}, & \text{if } i \neq j \end{cases},$$

where $\mathbf{D}$ represent the degree matrix expressed as follows

$$\mathbf{D}_{ij} := \begin{cases} \sum_{k=0}^{n} \mathbf{A}_{ik}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}. \tag{2.1}$$

On the other hand, in our analysis, we taking into account the normalized graph Laplacian $\mathbf{L}_N \in \mathbb{R}^{n \times n}$, defined as follows:

$$\mathbf{L}_N := \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}},$$

where $\mathbf{I}_n$ denotes the identity operator in $\mathbb{R}^n$. Since $\mathbf{A}$ is symmetric, also the graph Laplacian $\mathbf{L}_N$ is a symmetric matrix allowing us to compute its orthonormal eigendecomposition as

$$\mathbf{L}_N = \mathbf{U}\mathbf{M}_{\lambda}\mathbf{U}^{\mathrm{T}}, \tag{2.2}$$

where $\mathbf{M}_\lambda = \text{diag}(\lambda) = \text{diag}(\lambda_1, \ldots, \lambda_n)$ represents the diagonal matrix with eigenvalues λ_i, $i \in \{1, \ldots, n\}$ of $\mathbf{L}_N$ arranged increasingly along the diagonal entries. The columns $u_1, \ldots, u_n$ of the orthonormal matrix $\mathbf{U}$ represent the normalized eigenvectors of $\mathbf{L}_N$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. The ordered set $\hat{G} = \{u_1, \ldots, u_n\}$ of eigenvectors forms an orthonormal basis for the space of signals on the graph G. We denote $\hat{G}$ the spectrum of the graph G.
The result in (2.2) is obtained from the following propositions.

Proposition 2.1.1.
*If **A** is a real symmetric matrix, all of its eigenvalues are real numbers.*

Proof:
If $z = a + bi$ is a complex number, let $\bar{z} = a - bi$. Then $\overline{zw} = \bar{z}\bar{w}$, z is real if and only if $z = \bar{z}$, and $z \cdot \overline{z} = a^2 + b^2$. If λ is an eigenvalue of $\mathbf{A}$, for some $x \neq 0$, we have the following chain of equalities:

$$\begin{aligned}\lambda \bar{x}^{\mathrm{T}} x = &\bar{x}^{\mathrm{T}}(\lambda x) = \bar{x}^{\mathrm{T}}\mathbf{A}x = \\ &(\mathbf{A}^{\mathrm{T}}\bar{x})^{\mathrm{T}}x = (\mathbf{A}\bar{x})^{\mathrm{T}}x = \\ &(\bar{\mathbf{A}}\bar{x})^{\mathrm{T}}x = (\bar{\lambda}\bar{x})^{\mathrm{T}}x = \bar{\lambda}\bar{x}^{\mathrm{T}}x.\end{aligned}$$

Since $x \neq 0$, then $\bar{x}^{\mathrm{T}}x \neq 0$ and $\lambda = \bar{\lambda}$. □

Proposition 2.1.2.
*if **A** is a real symmetric matrix, its eigenvectors corresponding to different eigenvalues are orthogonal.*

Proof:
Suppose there are two eigenvectors, denoted as x and y, paired with their respective eigenvalues, λ and μ

$$\mathbf{A}x = \lambda x \quad \text{and} \quad \mathbf{A}y = \mu y, \lambda \neq \mu.$$

Then we have the following chain of equalities:

$$\lambda x^{\mathrm{T}}y = (\lambda x)^{\mathrm{T}}y = (\mathbf{A}x)^{\mathrm{T}}y = x^{\mathrm{T}}\mathbf{A}y = x^{\mathrm{T}}(\mu y) = \mu x^{\mathrm{T}}y.$$

Since $\lambda \neq \mu$, the above equalities implies that $x^{\mathrm{T}}y = 0$, that is $x \perp y$. □

Proposition 2.1.3.
If $\mathbf{A}$ *is a symmetric matrix, then* $\mathbf{A}$ *has n distinct eigenvectors that form an orthonormal basis for* $\mathbb{R}^n$*.*

Proof:
If every eigenvalues of $\mathbf{A}$ are distinct, then we are done, as indicated by the previous proposition. Otherwise, we will iteratively construct the eigenvectors as follows. Let u_1 be a normalized (i.e., re-scaled so that its norm is 1) eigenvector of $\mathbf{A}$ with corresponding eigenvalue λ_1. Assume we have k mutually orthogonal normalized eigenvectors $u_1, \ldots, u_k$, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$. We will now show how to construct a new eigenvector u_{k+1} with the eigenvalue λ_{k+1}, such that u_{k+1} is orthogonal to each of the vectors $u_1, \ldots, u_k$.
Let $\mathbf{U} = [u_1, \ldots, u_k] \in \mathbb{R}^{n\times k}$. Then $\mathbf{AU} = [\lambda_1 u_1, \ldots, \lambda_k u_k]$.
Let $\mathbf{V} = [v_{k+1}, \ldots, v_n] \in \mathbb{R}^{n\times(n-k)}$ be a matrix composed of any $n-k$ mutually orthogonal vectors such that the n vectors $u_1, \ldots, u_k, v_{k+1}, \ldots, v_n$ form an orthonormal basis for $\mathbb{R}^n$. Then note that

$$\mathbf{U}^{\mathrm{T}}\mathbf{V} = 0$$

and

$$\mathbf{V}^{\mathrm{T}}\mathbf{AU} = \mathbf{V}^{\mathrm{T}}[\lambda_1 u_1, \ldots, \lambda_k u_k] = 0.$$

Let w be an eigenvector of $\mathbf{V}^{\mathrm{T}}\mathbf{AV} \in \mathbb{R}^{(n-k)\times(n-k)}$ for some eigenvalue λ, so that $\mathbf{V}^{\mathrm{T}}\mathbf{AV} = \lambda w$. Then define $u_{k+1} = \mathbf{V}w$, and assume that w is rescaled if necessary so that $\|u_{k+1}\| = 1$. We now claim the following two statements are true:

(i) $\mathbf{U}^{\mathrm{T}}u_{k+1} = 0$, so that u_{k+1} is orthogonal to all of the columns of $\mathbf{U}$,

(ii) u_{k+1} is an eigenvector of $\mathbf{A}$, and λ is the corresponding eigenvalue of $\mathbf{A}$.

Note that, if **(i)** and **(ii)** are true, we can keep adding orthogonal vectors until $k = n$, completing the proof of the proposition.
To prove **(i)**, simply note that $\mathbf{U}^{\mathrm{T}}u_{k+1} = \mathbf{U}^{\mathrm{T}}\mathbf{V}w = 0w = 0$. To prove **(ii)**, let $d = \mathbf{A}u_{k+1} - \lambda u_{k+1}$. We need to show that $d = 0$. Note that $d = \mathbf{AV}w - \lambda\mathbf{V}w$, and so, $\mathbf{V}^{\mathrm{T}}d = \mathbf{V}^{\mathrm{T}}\mathbf{AV}w - \lambda\mathbf{V}^{\mathrm{T}}\mathbf{V}w = \mathbf{V}^{\mathrm{T}}\mathbf{AV}w - \lambda w = \lambda w - \lambda w = 0$. Therefore, $d = \mathbf{U}r$ for some

$r \in \mathbb{R}^k$, and then

$$r = \mathbf{U}^{\mathrm{T}}\mathbf{U}r = \mathbf{U}^{\mathrm{T}}d = \mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{V}w - \lambda\mathbf{U}^{\mathrm{T}}\mathbf{V}w = 0 - 0 = 0.$$

Thus $d = 0$, which completes the proof. □

Proposition 2.1.4.
If $\boldsymbol{A}$ is symmetric, then it can be decomposed as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$, where $\boldsymbol{U}$ is an orthonormal matrix and $\boldsymbol{D}$ is a diagonal matrix. The columns of $\boldsymbol{U}$ form an orthonormal basis of eigenvectors of $\boldsymbol{A}$, and the diagonal matrix $\boldsymbol{D}$ consists of the corresponding eigenvalues of $\boldsymbol{A}$.

Proof:
From Propositions (2.1.2) and (2.1.3), let $\mathbf{U} = [u_1, \ldots, u_n]$, where $u_1, \ldots, u_n$ are the n orthonormal eigenvectors of $\mathbf{A}$, and let

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

where $\lambda_1, \ldots, \lambda_n$ are the real corresponding eigenvalues as stated in Proposition (2.1.1). Then

$$(\mathbf{U}^{\mathrm{T}}\mathbf{U})_{ij} = u_i^{\mathrm{T}}u_j = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases},$$

whereby $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$, that is $\mathbf{U}^{\mathrm{T}} = \mathbf{U}^{-1}$.
Note that $\lambda_i\mathbf{U}^{\mathrm{T}}u_i = \lambda_i e_i$, with $i = 1, \ldots, n$, where e_i is the i-th unit vector.
Therefore:

$$\begin{aligned} \mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U} = \mathbf{U}^{\mathrm{T}}\mathbf{A}[u_1, \ldots, u_n] = \mathbf{U}^{\mathrm{T}}[\lambda_1 u_1, \ldots, \lambda_n u_n] = \\ = [\lambda_1 e_1, \ldots, \lambda_n e_n] = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = D. \end{aligned}$$

Thus $\mathbf{A} = (\mathbf{U}^{\mathrm{T}})^{-1}\mathbf{D}\mathbf{U}^{-1} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$. □

2.2 Fourier transform on graph

In classical Fourier analysis, such as in Euclidean space, the Fourier transform can be expressed in terms of the eigenvalues and eigenfunctions of the Laplace operator

$$\hat{x}(\omega) := \int_{\mathbb{R}} x(t) e^{-2\pi i t \omega} dt.$$

Here, the functions $u_\omega(t) = e^{-2\pi i t \omega}$ are the eigenfunctions of the Laplace operator $-\Delta = -\frac{d^2}{dt^2}$, i.e.,

$$-\Delta u_\omega(t) = -4\pi^2 i^2 \omega^2 u_\omega(t).$$

In addition, the signal $x(t)$ can be recovered from the Fourier coefficients $\hat{x}(\omega)$ using the inverse Fourier transform:

$$x(t) := \int_{\mathbb{R}} \hat{x}(\omega) e^{2\pi i t \omega} d\omega.$$

Similarly, we regard the elements of $\hat{G}$, specifically the eigenvectors $\{u_1, \ldots, u_n\}$, as the Fourier basis on the graph G. In particular, returning to our spatial signal x, we can define the graph Fourier transform of x as

$$\hat{x} := \mathbf{U}^{\mathrm{T}} x = (u_1^{\mathrm{T}} x, \ldots, u_n^{\mathrm{T}} x)^{\mathrm{T}}$$

and its inverse graph Fourier transform as

$$x := \mathbf{U}\hat{x}.$$

The frequency components or coefficients of the signal x with respect to the basis function u_i are represented by the entries $\hat{x}_i = u_i^{\mathrm{T}} x$ of $\hat{x}$. Therefore, $\hat{x} : \hat{G} \to \mathbb{R}$ can be viewed as a function on the spectral domain $\hat{G}$ of the graph G. To maintain the notation simple, we will also represent the spectral distribution $\hat{x}$ as vectors $(\hat{x}_1, \ldots, \hat{x}_n)$ in $\mathbb{R}^n$. Concerning the eigenvalues of the normalized graph Laplacian $\mathbf{L}_N$, it is well-know that:

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \leq 2. \tag{2.3}$$

It is worth noting that it is possible to use the spectral decomposition of other suitable operators on $\mathcal{L}(G)$ instead of the normalized graph Laplacian $\mathbf{L}_N$ to define the graph Fourier transform $\mathbf{U}$ on G. Common examples in the literature include the adjacency matrix $\mathbf{A}$ or other normalizations of $\mathbf{L}$. All the findings of this study hold true for these alternative generators of the graph Fourier transform as long as the Fourier matrix $\mathbf{U}$ remains orthogonal.
To establish the result of (2.3) and the previous consideration regarding the potential use of spectral decomposition of alternative operators on $\mathcal{L}(G)$, we present the following theorem:

Theorem 2.2.1.
1) For the normalized graph Laplacian $\boldsymbol{L}_N \in \mathbb{R}^n$ we have $\boldsymbol{L}_N$ is symmetric and

$$x^T \boldsymbol{L}_N x = \frac{1}{2} \sum_{(v_i,v_j)\in E} \boldsymbol{A}_{ij} \left(\frac{x(v_i)}{\boldsymbol{D}_{ii}^{1/2}} - \frac{x(v_j)}{\boldsymbol{D}_{jj}^{1/2}} \right)^2 . \tag{2.4}$$

This implies that $\boldsymbol{L}_N$ is positive semidefinite and its eigenvalues are all non-negative. Moreover, $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \leq 2$.
2) The eigenvalues of $\boldsymbol{L}_{RW}$ correspond to the eigenvalues of $\boldsymbol{L}_N$ and

- *(a)* $\sum_{k=1}^{n} \lambda_i = n$
- *(b) $\boldsymbol{L}_N$ has an eigenvalue $\lambda_1 = 0$ with the eigenvector $\boldsymbol{D}^{1/2}e$, the respective eigenvector $\boldsymbol{L}_{RW}$ for $\lambda_1 = 0$ is e*
- *(c) for $n \geq 0$ we have $\lambda_2 \leq \frac{n}{n-1}$ and $\lambda_n \geq \frac{n}{n-1}$*
- *(d)* $\lambda_i \in [0, 2]$ *for all* $i \in \{1, \ldots, n\}$

Before delving into the proof of this theorem, let's revisit the definition of the stochastic matrix $\mathbf{S}$ along with its properties, and also recall the definition of the random walk graph Laplacian $\mathbf{L}_{RW}$.
A matrix $\mathbf{S} \in \mathbb{R}^{n\times n}$ is called stochastic if all its entries $\mathbf{S}_{ij} \geq 0$ are non-negative and each column (row) sums to one, i.e., $\sum_{j=1}^{n} \mathbf{S}_{ij} = 1$ ($\sum_{i=1}^{n} \mathbf{S}_{ij} = 1$) for all $j \in \{1, \ldots, n\}$. In addition a stochastic matrix have the following properties:

- For the vector $e = (1, \ldots, 1)^{\mathrm{T}}$ we have $e^{\mathrm{T}}\mathbf{S} = e^{\mathrm{T}}$, i.e., e is a left eigenvector of $\mathbf{S}$ with respect to the eigenvalue $\lambda = 1$.
- In particular, $\lambda = 1$ is an eigenvalue of $\mathbf{S}$.

- For the column (row) sum 1-norm of the matrix **S**, we have

$$\|\mathbf{S}\|_1 = \max_i \sum_{j=1}^{n} |\mathbf{S}_{ij}| = 1 \ \ (\max_i \sum_{i=1}^{n} |\mathbf{S}_{ij}| = 1).$$

- For an arbitrary eigenpair (λ, x) of **S** we have

$$|\lambda|\|x\|_1 = \|\lambda x\|_1 = \|\mathbf{S}x\|_1 \leq \|\mathbf{S}\|_1\|x\|_1 = \|x\|_1. \tag{2.5}$$

  Therefore, all eigenvalues of **S** satisfy $|\lambda| \leq 1$.

The random walk graph Laplacian $\mathbf{L}_{RW} \in \mathbb{R}^{n\times n}$ is defined as:

$$\mathbf{L}_{RW} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1}\mathbf{A} = \mathbf{I}_n - \mathbf{S}^{\mathrm{T}}, \tag{2.6}$$

where $\mathbf{S}=\mathbf{A}^{\mathrm{T}}\mathbf{D}^{-1}$ represents the transition matrix in $\mathbb{R}^{n\times n}$. It is important to highlight that the matrix **S** is stochastic, which implies that the sum of the entries of its columns is one.

Proof of Theorem (2.2.1)**:**

1) To begin, we show that the normalized Laplacian $\mathbf{L}_N$ is symmetric. Given that **A** is symmetric by hypothesis and $\mathbf{I}_n$, **D** are diagonal matrices, resulting in $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$, $\mathbf{I}_n = {\mathbf{I}_n}^{\mathrm{T}}$, and $\mathbf{D} = \mathbf{D}^{\mathrm{T}}$. Then, we can establish the following sequence of equalities:

$$\begin{aligned}
\mathbf{L}_N^{\mathrm{T}} &= (\mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}})^{\mathrm{T}} = \\
&= \mathbf{I}_n^{\mathrm{T}} - (\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}})^{\mathrm{T}} = \\
&= \mathbf{I}_n - (\mathbf{D}^{\mathrm{T}})^{-1/2}\mathbf{A}^{\mathrm{T}}(\mathbf{D}^{\mathrm{T}})^{-1/2} = \\
&= \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} = \mathbf{L}_N.
\end{aligned}$$

To demonstrate the result in (2.4), we proceed as follows

$$\begin{aligned}
\sum_{(v_i,v_j)\in E} \mathbf{A}_{ij}\left(\frac{x(v_i)}{\mathbf{D}_{ii}^{1/2}} - \frac{x(v_j)}{\mathbf{D}_{jj}^{1/2}}\right)^2 &= \sum_{i,j=1}^{n} \mathbf{A}_{ij}\left(\frac{x^2(v_i)}{\mathbf{D}_{ii}} + \frac{x^2(v_j)}{\mathbf{D}_{jj}} - \frac{2x(v_i)x(v_j)}{\sqrt{\mathbf{D}_{ii}\mathbf{D}_{jj}}}\right) = \\
= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} \mathbf{A}_{ij}\frac{x^2(v_i)}{\mathbf{D}_{ii}}\right) &+ \sum_{j=1}^{n}\left(\sum_{i=1}^{n} \mathbf{A}_{ji}\frac{x^2(v_j)}{\mathbf{D}_{jj}}\right) - 2\sum_{i,j=1}^{n} \mathbf{A}_{ij}\frac{x(v_i)x(v_j)}{\sqrt{\mathbf{D}_{ii}\mathbf{D}_{jj}}} = \\
= \sum_{i=1}^{n} \frac{\mathbf{D}_{ii}x^2(v_i)}{\mathbf{D}_{ii}} &+ \sum_{j=1}^{n} \frac{\mathbf{D}_{jj}x^2(v_j)}{\mathbf{D}_{jj}} - 2\sum_{i,j=1}^{n} \mathbf{A}_{ij}\frac{x(v_i)x(v_j)}{\sqrt{\mathbf{D}_{ii}\mathbf{D}_{jj}}} =
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \frac{\mathbf{D}_{ii} x^2(v_i)}{\mathbf{D}_{ii}} + \sum_{j=1}^{n} \frac{\mathbf{D}_{jj} x^2(v_j)}{\mathbf{D}_{jj}} - 2 \sum_{i,j=1}^{n} \mathbf{A}_{ij} \frac{x(v_i) x(v_j)}{\sqrt{\mathbf{D}_{ii} \mathbf{D}_{jj}}} = \\
&= \sum_{i=1}^{n} \frac{\mathbf{D}_{ii} x^2(v_i)}{\mathbf{D}_{ii}} + \sum_{j=1}^{n} \frac{\mathbf{D}_{jj} x^2(v_j)}{\mathbf{D}_{jj}} - 2 \sum_{i,j=1}^{n} \mathbf{A}_{ij} \frac{x(v_i) x(v_j)}{\sqrt{\mathbf{D}_{ii} \mathbf{D}_{jj}}} = \\
&= x^{\mathrm{T}} \mathbf{I}_n x - 2x^{\mathrm{T}} \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} x + x^{\mathrm{T}} \mathbf{I}_n x = \\
&= 2x^{\mathrm{T}} \mathbf{I}_n x - 2x^{\mathrm{T}} \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} x = \\
&= 2x^{\mathrm{T}} \left(\mathbf{I}_n - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} \right) x = 2x^{\mathrm{T}} \mathbf{L}_N x \Rightarrow \\
&\Rightarrow x^{\mathrm{T}} \mathbf{L}_N x = \frac{1}{2} \sum_{(\mathrm{v}_i, \mathrm{v}_j) \in E} \mathbf{A}_{ij} \left(\frac{x(\mathrm{v}_i)}{\mathbf{D}_{ii}^{1/2}} - \frac{x(\mathrm{v}_j)}{\mathbf{D}_{jj}^{1/2}} \right)^2 .
\end{aligned}
$$

2) Based on the previous point, we know that the normalized Laplacian $\mathbf{L}_N$ is symmetric, positive semidefinite and its eigenvalues are all non-negative. To ensure the same eigenvalues between the matrices, $\mathbf{L}_N$ and $\mathbf{L}_{RW}$, they should be similar. We recall that two matrices, $\mathbf{Q}$ and $\mathbf{P}$ in $\mathbb{R}^{n \times n}$, are similar if there exists an invertible matrix $\mathbf{U}$ in $\mathbb{R}^{n \times n}$ such that $\mathbf{P} = \mathbf{U}^{-1}\mathbf{Q}\mathbf{U}$. Therefore, we can observe that $\mathbf{L}_N = \mathbf{D}^{1/2}\mathbf{L}_{RW}\mathbf{D}^{-1/2} \Rightarrow \mathbf{L}_N$ and $\mathbf{L}_{RW}$ are similar $\Rightarrow \mathbf{L}_N$ and $\mathbf{L}_{RW}$ have the same eigenvalues.
Let u_k represent the eigenvector of $\mathbf{L}_N$ corresponding to the eigenvalue λ_k, then $v_k = \mathbf{D}^{-1/2} u_k$ is the eigenvector of $\mathbf{L}_{RW}$ corresponding to the same eigenvalue λ_k:

$$
\begin{aligned}
\mathbf{L}_{RW} v_k &= \left(\mathbf{I}_n - \mathbf{D}^{-1} \mathbf{A} \right) v_k = \\
&= \mathbf{D}^{-1/2} \left(\mathbf{D}^{1/2} - \mathbf{D}^{-1/2} \mathbf{A} \right) \mathbf{D}^{-1/2} u_k = \\
&= \mathbf{D}^{-1/2} \left(\mathbf{I}_n - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} \right) u_k = \\
&= \mathbf{D}^{-1/2} \mathbf{L}_N u_k = \mathbf{D}^{-1/2} \lambda_k u_k = \lambda_k \mathbf{D}^{-1/2} u_k = \lambda_k v_k .
\end{aligned}
$$

(a) We can observe the following:

$$
\sum_{k=1}^{n} \lambda_k = \mathrm{tr}\left(\mathbf{L}_N \right) = \mathrm{tr} \left(\begin{bmatrix} 1 & * & \dots & * \\ * & 1 & * & \vdots \\ \vdots & * & \ddots & * \\ * & \dots & * & 1 \end{bmatrix} \right) = n, \text{ with } \lambda_1 = 0.
$$

To ensure the first equality, let's recall the definition of the characteristic polynomial of a matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$: $p(t) = \det(\mathbf{P} - t\mathbf{I})(-1)^n \left[t^n - \mathrm{tr}(\mathbf{P}) t^{n-1} + \ldots + (-1)^n \det(\mathbf{P}) \right]$. On the other

hand, $p(t) = (-1)^n(t-\lambda_1)\dots(t-\lambda_n)$, where λ_k for $k \in \{1,\dots,n\}$ are the eigenvalues of the square matrix **P**. Therefore, by comparing coefficients, we find that tr(**P**)$= \lambda_1 + \dots + \lambda_n$.
(b) We can proceed as follows:

$$\mathbf{L}_{RW}e = \left(\mathbf{I}_n - \mathbf{D}^{-1}\mathbf{A}\right)e = \mathbf{D}^{-1}\left(\mathbf{D}-\mathbf{A}\right)e = \mathbf{D}^{-1}\mathbf{L}e = \mathbf{D}^{-1}0 = 0$$

$\Rightarrow$ e is the eigenvector of $\mathbf{L}_{RW}$ corresponding to the eigenvalue $\lambda_1 = 0 \Rightarrow \mathbf{D}^{1/2}e$ is the respective eigenvector of $\mathbf{L}_N$ corresponding to the eigenvalue $\lambda_1 = 0$.
(c) Given that the eigenvalues are arranged in increasing order, $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$, we can obtain the following:

$$n = \lambda_2 + \dots + \lambda_n \geq \lambda_2 + (n-2)\lambda_2 = (n-1)\lambda_2 \Rightarrow \lambda_2 \leq \frac{n}{n-1}$$
$$n = \lambda_2 + \dots + \lambda_n \leq (n-2)\lambda_n + \lambda_n = (n-1)\lambda_n \Rightarrow \lambda_n \geq \frac{n}{n-1}$$

(d) We recall the definition of the random walk graph Laplacian, denoted as $\mathbf{L}_{RW} = \mathbf{I}_n - \mathbf{S}^\mathrm{T}$, where **S** is a stochastic matrix. With the result provided in (2.5) for stochastic matrices, we can infer that all the eigenvalues of $\mathbf{S}^\mathrm{T}$ are between -1 and $1 \Rightarrow$ all the eigenvalues of $\mathbf{L}_{RW}$ are between 0 and 2. □

2.3 Convolution on graphs

The graph Fourier transform enables the definition of a convolution operation between two graph signals, x and y. In classical Fourier analysis, the convolution of two signals, in time domain, is computed as follows:

$$(x * y)(t) = \int_{\mathbb{R}} x(s)y(t-s)\,ds.$$

By the convolution theorem, its Fourier transform is:

$$\widehat{x * y}(\omega) = \widehat{x}(\omega) \cdot \widehat{y}(\omega),$$

where $\hat{x}(\omega)$ and $\hat{y}(\omega)$ are the Fourier transforms of $x(t)$ and $y(t)$, respectively.
In particular, we define for $x, y \in \mathcal{L}(G)$ the graph convolution as

$$x * y := \mathbf{U}(\mathbf{M}_{\hat{x}}\hat{y}) = \mathbf{U}\mathbf{M}_{\hat{x}}\mathbf{U}^{\mathrm{T}}y. \tag{2.7}$$

As before, $\mathbf{M}_{\hat{x}}$ represents the diagonal matrix defined as $\mathbf{M}_{\hat{x}} = \text{diag}(\hat{x})$. Moreover, $\mathbf{M}_{\hat{x}}\hat{y} = (\hat{x}_1\hat{y}_1, \ldots, \hat{x}_n\hat{y}_n)$ yields the point-wise product of the two vectors $\hat{x}$ and $\hat{y}$.

Proposition 2.3.1.
The convolution operator $*$ on $\mathcal{L}(G)$ has the following properties:

(1) Commutativity: $x * y = y * x$

(2) Associativity: $(x * y) * z = x * (y * z)$

(3) Distributivity: $(x + y) * z = x * z + y * z$

(4) Associativity for scalar multiplication: $(\alpha x) * y = \alpha(y * x)$ *for all* $\alpha \in \mathbb{R}$

Proof:
(1) Given that $\hat{x}\hat{y} = \text{diag}(\hat{x})\hat{y} = \text{diag}(\hat{y})\hat{x}$, we have:

$$x * y = \mathbf{U}\text{diag}(\hat{x})\hat{y} = \mathbf{U}\text{diag}(\hat{y})\hat{x} = y * x$$

(2) Due to point **(1)**, we obtain the following:

$$\begin{aligned}(x * y) * z &= (\mathbf{U}\text{diag}(\hat{x})\hat{y}) * z = \\ &= z * (\mathbf{U}\text{diag}(\hat{x})) = \mathbf{U}\text{diag}(\hat{z})(\widehat{\mathbf{U}\text{diag}(\hat{x})}\hat{y}) = \\ &= \mathbf{U}\text{diag}(\hat{z})\mathbf{U}^{\mathrm{T}}\mathbf{U}\text{diag}(\hat{x})\hat{y} = \mathbf{U}\text{diag}(\hat{x})\text{diag}(\hat{z})\hat{y} = \\ &= \mathbf{U}\text{diag}(\hat{x})\mathbf{U}^{\mathrm{T}}\mathbf{U}\text{diag}(\hat{y})\hat{z} = \mathbf{U}\text{diag}(\hat{x})(\widehat{\mathbf{U}\text{diag}(\hat{y})}\hat{z}) = \\ &= x * (\mathbf{U}\text{diag}(\hat{y})\hat{z}) = x * (y * z)\end{aligned}$$

(3) As $\widehat{x + y} = \mathbf{U}^{\mathrm{T}}(x + y) = \mathbf{U}^{\mathrm{T}}x + \mathbf{U}^{\mathrm{T}}y = \hat{x} + \hat{y}$, we have the following chain of equalities:

$$\begin{aligned}(x + y) * z &= \mathbf{U}\text{diag}(\widehat{x + y})\hat{z} = \\ &= \mathbf{U}\text{diag}(\hat{x} + \hat{y})\hat{z} = \\ &= \mathbf{U}[\text{diag}(\hat{x}) + \text{diag}(\hat{y})]\hat{z} = \\ &= \mathbf{U}\text{diag}(\hat{x})\hat{z} + \mathbf{U}\text{diag}(\hat{y})\hat{z} = x * z + y * z\end{aligned}$$

(4) Since $\widehat{\alpha x} = \mathbf{U}^{\mathrm{T}} \alpha x = \alpha \mathbf{U}^{\mathrm{T}} x = \alpha \hat{x}$ for $\alpha \in \mathbb{R}$, we have the following chain of equalities:

$$\begin{aligned}(\alpha x) * y &= \mathbf{U}\mathrm{diag}(\widehat{\alpha x})\hat{y} = \\ &= \mathbf{U}\mathrm{diag}(\alpha \hat{x})\hat{y} = \\ &= \alpha \mathbf{U}\mathrm{diag}(\hat{x})\hat{y} = \alpha(x * y)\end{aligned}$$

Therefore, all the properties are satisfied. □

The unity element of the convolution is given by $f_{\mathbb{1}} = \sum_{i=1}^{n} u_i$. Considering the linear structure in (2.7), we can further define a convolution operator $\mathbf{C}_x$ on $\mathcal{L}(G)$ as:

$$\mathbf{C}_x = \mathbf{U}\mathbf{M}_{\hat{x}}\mathbf{U}^{\mathrm{T}}.$$

The convolution $x * y$ can then be formulated as the matrix-vector product $\mathbf{C}_x y = x * y$. Expressed in this way, each signal $x \in \mathcal{L}(G)$ can be interpreted as a filter function that operates through convolution on a second signal y.

2.4 The graph C^*-algebra

The properties **(1)**-**(4)** of Proposition (2.3.1) of graph convolution ensure that the vector space $\mathcal{L}(G)$, equipped with the convolution operation $*$, operates as a commutative and associative algebra. Considering the identity as a trivial involution and the norm

$$\|x\|_{\mathcal{A}} = \sup_{\|y\| = 1} \|x * y\|,$$

we obtain a real C^*-algebra $\mathcal{A}$. Significant for us is the fact that $\mathcal{A}$ is both commutative and finite. A general introduction to C^*-algebras, including discussions on commutative and finite C^*-algebras, is available in [28]. In addition, further insights about the characterization of real C^*-algebras can be found in [29].

In this study, the graph C^*-algebra $\mathcal{A}$ operates as the standard model for graph signal processing. The algebra $\mathcal{A}$ contains all potential signals and filter functions on G, illustrating how filters interact with signals through convolution. Moreover, $\mathcal{A}$ includes the complete information of the graph Fourier transform. This can be viewed as follows: the spectrum of the commutative C^*-algebra $\mathcal{A}$ consists of multiplicative linear functionals on $\mathcal{A}$ that maintain the algebra's multiplicative structure. In our scenario, these characters of $\mathcal{A}$ precisely correspond to the n functionals $x \to u_k^\mathrm{T} x$, where $k \in \{1, \ldots, n\}$. Hence, the spectrum of the C^*-algebra $\mathcal{A}$ can be naturally associated with the previously introduced spectrum $\hat{G} = \{u_1, \ldots, u_n\}$ of the graph G. When applied to the C^*-algebra $\mathcal{A}$, the renowned Gelfand-Naimark Theorem states that the graph Fourier transform $\mathbf{U}^\mathrm{T}$ operates as an algebra isomorphism between the C^*-algebra $\mathcal{A}$ and the C^*-algebra of functions on the spectrum $\hat{G}$, employing point-wise multiplication as the multiplicative operation. Our investigation of positive definite functions on the graph G is essentially based on this algebraic signals model.

The algebraic structure for graph signal processing has been explored in several previous work. In [30], a comprehensive algebraic signal model was formulated to describe signal processing in discrete environments. The C^*-algebra framework considered in this study corresponds as a particular instance within the general scenario in [30]. However, it establishes a significantly closer relationship among signals, convolution, and the graph Fourier transform. From now on, to keep the notation simple, we denote the normalized graph Laplacian $\mathbf{L}_N$ as $\mathbf{L}$, unless otherwise specified.

Besides the C^*-algebra $\mathcal{A}$ we will consider the following two subalgebras:

(1) The C^*-algebra $\mathcal{A}_\mathbf{L}$ is generated from the normalized Laplacian graph $\mathbf{L}$ as

$$\mathcal{A}_\mathbf{L} := \operatorname{span}\left\{f_\mathbb{1}, \mathbf{L} f_\mathbb{1}, \mathbf{L}^2 f_\mathbb{1}, \ldots, \mathbf{L}^{n-1} f_\mathbb{1}\right\}.$$

$\mathcal{A}_\mathbf{L}$ is a subalgebra of $\mathcal{A}$ that contains the unity element $f_\mathbb{1}$ of the convolution. The algebra $\mathcal{A}_\mathbf{L}$ is significant for the construction of filter functions based on the normalized Laplacian $\mathbf{L}$.

(2) The C^*-algebra $\mathcal{B}_M$, representing bandlimited signals with a bandwidth of $M \leq n$, is defined as

$$\mathcal{B}_M := \operatorname{span}\{u_1, \ldots, u_M\}.$$

If $M < n$, then $f_\mathbb{1}$ does not belong to $\mathcal{B}_M$. In this scenario, the multiplicative unity of the subalgebra $\mathcal{B}_M$ is defined as $\sum_{k=1}^M u_k$. Bandlimited signals are significant for us as approximation spaces.

We conclude this section by providing a characterization of the subalgebra $\mathcal{A}_\mathbf{L}$.

Proposition 2.4.1.

Assume that the normalized graph Laplacian $\mathbf{L}$ *has precisely* $r \leq n$ *distinct eigenvalues. Then,* $\mathcal{A}_L$ *is r-dimensional subalgebra of* $\mathcal{A}$. *A signal* x *is contained in* $\mathcal{A}_L$ *if and only if* $\hat{x}_k = \hat{x}_{k'}$

whenever $\lambda_k = \lambda_{k'}$. *Furthermore,*

$$\mathcal{A}_L = \left\{ f_{\mathbb{1}}, \boldsymbol{L} f_{\mathbb{1}}, \ldots, \boldsymbol{L}^{r-1} f_{\mathbb{1}} \right\}.$$

Proof:
We consider the C^*-algebra

$$\tilde{\mathcal{A}} = \{ x \in \mathcal{A} \mid \hat{x}_k = \hat{x}_{k'} \text{ if } \lambda_k = \lambda_{k'} \},$$

and show that $\tilde{\mathcal{A}}$ correspond to $\mathcal{A}_{\mathrm{L}}$. We start by showing that $\tilde{\mathcal{A}}$ is a C^*-subalgebra of $\mathcal{A}$. Indeed, $\tilde{\mathcal{A}}$ is closed under **(a)** addition, **(b)** scalar multiplication and **(c)** convolution operator:

(a) Closure under addition: let $x, y \in \tilde{\mathcal{A}}$, we need to show that $x + y \in \tilde{\mathcal{A}}$. Given x and y belonging to $\tilde{\mathcal{A}}$, it follows that $\hat{x}_k = \hat{x}_{k'}$ and $\hat{y}_k = \hat{y}_{k'}$ if $\lambda_k = \lambda_{k'}$. Therefore, $(\hat{x}+\hat{y})_k = \hat{x}_k + \hat{y}_k = \hat{x}_{k'} + \hat{y}_{k'} = (\hat{x}+\hat{y})_{k'}$ if $\lambda_k = \lambda_{k'}$, which implies that $x + y \in \tilde{\mathcal{A}}$

(b) Closure under scalar multiplication: let $x \in \tilde{\mathcal{A}}$ and $\alpha \in \mathbb{R}$, we need to show that $\alpha x \in \tilde{\mathcal{A}}$. Since x belongs to $\tilde{\mathcal{A}}$, it follows that $\hat{x}_k = \hat{x}_{k'}$ if $\lambda_k = \lambda_{k'}$. Then, $\alpha \hat{x}_k = \alpha \hat{x}_{k'}$ if $\lambda_k = \lambda_{k'}$, which implies that $\alpha x \in \tilde{\mathcal{A}}$

(c) Closure under convolution opeartor: let $x, y \in \tilde{\mathcal{A}}$, we need to show that $(x * y) \in \tilde{\mathcal{A}}$. Given x and y belonging to $\tilde{\mathcal{A}}$, it follows that $\hat{x}_k = \hat{x}_{k'}$ and $\hat{y}_k = \hat{y}_{k'}$ if $\lambda_k = \lambda_{k'}$. Thus:

$$\begin{aligned}(\widehat{x * y})_k &= (\mathbf{U}^{\mathrm{T}} \mathbf{U} \mathbf{M}_{\hat{x}}(\hat{x}) \hat{y})_k = (\mathbf{M}_{\hat{x}} \hat{y})_k = \\ &= (\hat{x}_1 \hat{y}_1, \ldots, \hat{x}_n \hat{y}_n)_k = \hat{x}_k \hat{y}_k = \\ &= \hat{x}_{k'} \hat{y}_{k'} = (\hat{x}_1 \hat{y}_1, \ldots, \hat{x}_n \hat{y}_n)_{k'} = \\ &= (\mathbf{M}_{\hat{x}} \hat{y})_{k'} = (\mathbf{U}^{\mathrm{T}} \mathbf{U} \mathbf{M}_{\hat{x}}(\hat{x}) \hat{y})_{k'} = (\widehat{x * y})_{k'}\end{aligned}$$

which implies that $(\widehat{x * y})_k \in \tilde{\mathcal{A}}$.

Furthermore, the C^*-subalgebra $\tilde{\mathcal{A}}$ contains the unity element of the convolution, $f_{\mathbb{1}} = \sum_{i=1}^{n} u_i$. Therefore, we need to show the following: $(\hat{f}_{\mathbb{1}})_k = (\hat{f}_{\mathbb{1}})_{k'}$ if $\lambda_k = \lambda_{k'}$. To establish this claim, we can observe the following:

$$\hat{f}_{\mathbb{1}} = \mathbf{U}^{\mathrm{T}} \left(\sum_{i=1}^{n} u_i \right) = \left(u_1^{\mathrm{T}} \sum_{i=1}^{n} u_i, \ldots, u_n^{\mathrm{T}} \sum_{i=1}^{n} u_i \right) = (1, \ldots, 1).$$

Since $\hat{f}_{\mathbb{1}} = (1, \ldots, 1)$, it follows that $(\hat{f}_{\mathbb{1}})_k = (\hat{f}_{\mathbb{1}})_{k'}$ if $\lambda_k = \lambda_{k'}$ for all $k, k' \in \{1, \ldots, n\}$. Thus $f_{\mathbb{1}} \in \tilde{\mathcal{A}}$.

Given that $\mathbf{U}^\mathrm{T}\mathbf{L}x = \mathbf{M}_\lambda \mathbf{U}^\mathrm{T} x$, we can observe the following:

$$
\begin{aligned}
\mathbf{L}x \in \tilde{\mathcal{A}} \iff & \mathbf{L}x \in \mathcal{A} : (\widehat{\mathbf{L}x})_k = (\widehat{\mathbf{L}x})_{k'} \text{ if } \lambda_k = \lambda_{k'} \iff \\
\iff & \mathbf{L}x \in \mathcal{A} : (\mathbf{U}^\mathrm{T}\mathbf{L}x)_k = (\mathbf{U}^\mathrm{T}\mathbf{L}x)_{k'} \text{ if } \lambda_k = \lambda_{k'} \iff \\
\iff & x \in \mathcal{A} : (\mathbf{M}_\lambda \mathbf{U}^\mathrm{T}x)_k = (\mathbf{M}_\lambda \mathbf{U}^\mathrm{T}x)_{k'} \text{ if } \lambda_k = \lambda_{k'} \iff \\
\iff & x \in \mathcal{A} : (\mathbf{M}_\lambda \hat{x})_k = (\mathbf{M}_\lambda \hat{x})_{k'} \text{ if } \lambda_k = \lambda_{k'} \iff \\
\iff & x \in \mathcal{A} : \lambda_k \hat{x}_k = \lambda_{k'} \hat{x}_{k'} \text{ if } \lambda_k = \lambda_{k'} \iff \\
\iff & x \in \mathcal{A} : \hat{x}_k = \hat{x}_{k'} \text{ if } \lambda_k = \lambda_{k'} \iff x \in \tilde{\mathcal{A}},
\end{aligned}
$$

consequently, $\mathbf{L}x \in \tilde{\mathcal{A}}$ if $x \in \tilde{\mathcal{A}}$. This implies that $\mathcal{A}_\mathbf{L}$ is contained in $\tilde{\mathcal{A}}$. As the dimension of $\tilde{\mathcal{A}}$ corresponds to the number r of distinct eigenvalues of $\mathbf{L}$, we only need to show that $\{f_\mathbb{1}, \mathbf{L}f_\mathbb{1}, \dots, \mathbf{L}^{r-1}f_\mathbb{1}\} \subset \mathcal{A}_\mathbf{L}$ forms a system of r linearly independent vector space elements. To achieve this, we select r distinct eigenvalues of the normalized graph Laplacian $\mathbf{L}$ and denote them as $\lambda_{k_1} \dots, \lambda_{k_r}$. Then, we have:

$$
\begin{aligned}
u_{k_j}^\mathrm{T} \mathbf{L}^{i-1} f_\mathbb{1} &= (u_{k_j}^\mathrm{T}\mathbf{L})\mathbf{L}^{i-2} f_\mathbb{1} = \lambda_{k_j} u_{k_j}^\mathrm{T} \mathbf{L}^{i-2} f_\mathbb{1} = \\
&= \lambda_{k_j}^2 u_{k_j}^\mathrm{T} \mathbf{L}^{i-3} f_\mathbb{1} = \dots = \lambda_{k_j}^{i-1} u_{k_j}^\mathrm{T} f_\mathbb{1} = \\
&= \lambda_{k_j}^{i-1} u_{k_j}^\mathrm{T} \sum_{i=1}^{r} u_{k_i} = \lambda_{k_j}^{i-1} \sum_{i=1}^{r} \underbrace{u_{k_j}^\mathrm{T} u_{k_i}}_{\delta_{ij}} = \lambda_{k_j}^{i-1}, \text{ for } i, j \in \{1, \dots, r\}.
\end{aligned}
$$

Now, since the Vamdermonde matrix

$$
\mathbf{V}_r = \begin{pmatrix} \lambda_{k_1}^0 & \lambda_{k_1}^1 & \dots & \lambda_{k_1}^{r-1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k_r}^0 & \lambda_{k_r}^1 & \dots & \lambda_{k_r}^{r-1} \end{pmatrix}
$$

has the following determinant

$$
\det(\mathbf{V}_r) = \prod_{1 \le i < j \le r} (\lambda_{k_j} - \lambda_{k_i}) = (\lambda_{k_2} - \lambda_{k_1}) \cdots (\lambda_{k_r} - \lambda_{k_{r-1}}) \neq 0, \tag{2.8}
$$

is invertible. Thus the matrix $(f_\mathbb{1}, \mathbf{L}f_\mathbb{1}, \dots, \mathbf{L}^{r-1}f_\mathbb{1})$ has full rank and its columns are linearly independent. $\square$

Proposition (2.4.1) asserts that the equality $\mathcal{A} = \mathcal{A}_\mathbf{L}$ holds true if and only if the spectrum of the normalized graph Laplacian $\mathbf{L}$ is simple, meaning that all eigenvalues of $\mathbf{L}$ are distinct. In this scenario, the dimension of the C^*-algebra $\mathcal{A}_{\mathbf{L}}$ reaches its maximum capacity. However, for certain graphs, the C^*-subalgebra $\mathcal{A}_{\mathbf{L}}$ can also be particularly small: If G represents an unweighted complete graph, meaning every pair of nodes in V is linked by a unique, equally weighted edge, then the normalized graph Laplacian has only the eigenvalues $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = \ldots = \lambda_n = \frac{n}{n-1}$. In this case, $\mathcal{A}_{\mathbf{L}} = \text{span}\,\{f_{\mathbb{1}}, \mathbf{L}f_{\mathbb{1}}\}$ is only two-dimensional. In order to see the example, suppose we have a complete graph $\mathbf{K}_n$ with n vertices. Thus, the adjacency matrix $\mathbf{A}$ and the degree matrix $\mathbf{D}$ will each be an $n \times n$ matrix, where the elements $\mathbf{A}_{ij}$ and the elements $\mathbf{D}_{ij}$ are defined as follows:

$$\mathbf{A}_{ij} = \begin{cases} 1, & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases} \qquad \mathbf{D}_{ij} = \begin{cases} n-1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

Hence, the adjacency matrix and the degree matrix of the complete graph will be

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} n-1 & 0 & \cdots & 0 \\ 0 & n-1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n-1 \end{pmatrix}.$$

Consequently, the normalized graph Laplacian $\mathbf{L}$, defined as $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, is

$$\mathbf{L} = \begin{pmatrix} 1 & \frac{-1}{n-1} & \cdots & \frac{-1}{n-1} \\ \frac{-1}{n-1} & 1 & \cdots & \frac{-1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{n-1} & \frac{-1}{n-1} & \cdots & 1 \end{pmatrix}.$$

To verify that 0 is an eigenvalue with geometric multiplicity $m_g = 1$, we show that the dimension of the eigenspace associated with this eigenvalue is equal to one, i.e., $\dim(\text{Ker}\,(\mathbf{L} - 0\mathbf{I}_n))$

is 1:

$$\text{Ker}(\mathbf{L} - 0\mathbf{I}_n) = \text{Ker}\left(\begin{bmatrix} 1 & \frac{-1}{n-1} & \cdots & \frac{-1}{n-1} \\ \frac{-1}{n-1} & 1 & \cdots & \frac{-1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{n-1} & \frac{-1}{n-1} & \cdots & 1 \end{bmatrix}\right). \tag{2.9}$$

Then, the associated linear system consists of j equations: $x_j - \frac{1}{n-1}\sum_{i\neq j} x_i$ for all $i, j \in \{1, \dots, n\}$. Keeping the equation for $j = 1$ unchanged, we can obtain $n-1$ equations by subtracting the j-th equation from the $(j-1)$-th equation, for all $j \in \{2, \dots, n\}$:

$$x_j - x_{j-1} - \frac{1}{n-1}\underbrace{\left(\sum_{i\neq j} x_i - \sum_{i\neq j-1} x_i\right)}_{x_{j-1}-x_j} = 0 \iff \left(1 + \frac{1}{n-1}\right)(x_j - x_{j-1}) = 0 \iff$$
$$\iff x_j = x_{j-1}, \text{ for all } j \in \{2, \dots, n\}.$$

Therefore, we obtain the subsequent associated linear system:

$$\begin{cases} x_1 - \frac{1}{n-1}\sum_{i=2}^{n} x_i = 0 \\ x_j = x_{j-1}, \text{ for all } j \in \{2, \dots, n\} \end{cases} \iff \begin{cases} x_1 - \frac{n-1}{n-1} x_1 = 0 \\ x_j = x_{j-1}, \text{ for all } j \in \{2, \dots, n\} \end{cases} \iff$$

$$\iff \begin{cases} 0 = 0 \\ x_j = x_{j-1}, \text{ for all } j \in \{2, \dots, n\} \end{cases} \iff \begin{cases} x_2 = x_1 \\ \vdots \\ x_n = x_{n-1} \end{cases} \iff \begin{cases} x_2 = x_1 \\ \vdots \\ x_n = x_1 \end{cases}.$$

Currently, taking up the expression in (2.9), we can conclude:

$$\text{Ker}(\mathbf{L} - 0\mathbf{I}_n) = \left\{(x_1, \dots, x_1)^{\mathrm{T}} \mid x_1 \in \mathbb{R}\right\} = \left\langle \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\rangle \Rightarrow$$
$$\Rightarrow \dim[\text{Ker}\,(\mathbf{L} - 0\mathbf{I}_n)] = 1.$$

Now, we need to establish that $\frac{n}{n-1}$ is an eigenvalue with geometric multiplicity $m_g = n - 1$. We prove that the dimension of the eigenspace related to this eigenvalue is equal to $n-1$, that is, the $\dim(\text{Ker}(\mathbf{L} - \frac{n}{n-1}\mathbf{I}_n))$ is equal to $n-1$:

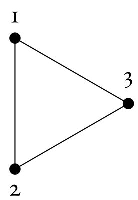
$$\text{Ker}\left(\mathbf{L} - \frac{n}{n-1}\mathbf{I}_n\right) = \text{Ker}\left(\begin{bmatrix} \frac{-1}{n-1} & \frac{-1}{n-1} & \cdots & \frac{-1}{n-1} \\ \frac{-1}{n-1} & \frac{-1}{n-1} & \cdots & \frac{-1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{n-1} & \frac{-1}{n-1} & \cdots & \frac{-1}{n-1} \end{bmatrix}\right) \sim \text{Ker}\left(\begin{bmatrix} 1 & \dots & 1 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}\right).$$

Thus, the associated linear system consists of only one equation: $x_1 + \ldots + x_n = 0 \Rightarrow x_1 = -x_2 - \ldots - x_n$. In conclusion, taking into account the above expression, we obtain:

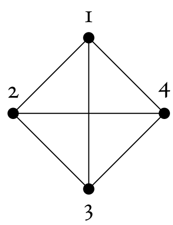
$$\text{Ker}\left(\mathbf{L} - \frac{n}{n-1}\mathbf{I}_n\right) = \left\{(-x_2 - \ldots - x_n, x_2, \ldots, x_n)^{\mathrm{T}} \mid (x_2, \ldots, x_n) \in \mathbb{R}^{n-1}\right\} =$$

$$= \langle \begin{pmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \rangle \Rightarrow \dim\left[\text{Ker}\left(\mathbf{L} - \frac{n}{n-1}\mathbf{I}_n\right)\right] = n - 1.$$

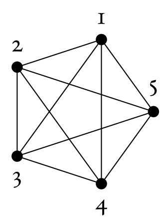
Unweighted complete graph $\mathbf{K}_n$



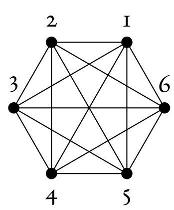
K_3



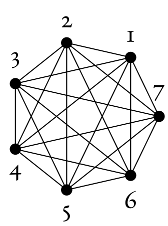
K_4



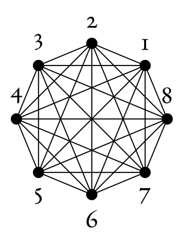
K_5



K_6



K_7



K_8

3

Kernel-Based Methods For Interpolation On Graphs

In this chapter, we provide a synthesis of well-know principles regarding kernel-based interpolation methods applied to discrete set. Within the traditional Euclidean framework, similar concepts can be found in references such as [16], [31] or in the comprehensive discussion provided in [17]. A general overview of kernel-based methodologies in machine learning is given in [32]. A recent investigation on the history and research trends related to positive definite kernels has been documented in [33]. It is important to highlight that subsequent derivations do not yet take into account the spectral or geometric information of the graph G.

3.1 Preliminaries

In this section we provide a series of notions related to kernels, which are essential for grasp the upcoming discussions.

Definition 3.1.1. ***(Norm)***
A norm is a function on a vector space V over $\mathbb{R}$ from V to $\mathbb{R}$ satisfying the following properties $\forall u, v \in V$ and $\forall \alpha \in \mathbb{R}$:

- ***(a)*** *Non-Negative:* $\|u\| \geq 0$

(b) *Strictly-Positive:* $\|u\| = 0 \Rightarrow u = 0$

(c) *Homogeneous:* $\|\alpha u\| = |\alpha|\|u\|$

(d) *Triangle Inequality:* $\|x + y\| \leq \|u\| + \|v\|$

Definition 3.1.2. ***(Inner Product)***
Given a vector space V over $\mathbb{R}$*, an* ***internal or scalar product*** *is a function* $\langle \cdot, \cdot \rangle$*:* $V \times V \to \mathbb{R}$*, which is a symmetric positive definite bilinear form:* $\forall u, v, w \in \mathcal{H}$ *(Hilbert Space),* $\alpha \in \mathbb{R}$

(a) *Symmetry:* $\langle v, w \rangle = \langle w, v \rangle$

(b) *Linearity with respect to first term:* $\langle v + w, u \rangle = \langle v, u \rangle + \langle w, u \rangle$

(c) *Linearity with respect to second term:* $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle$

(d) *Associative:* $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$

(e) *Positive Definite:* $\langle v, v \rangle > 0 \, \forall v \neq 0$

Definition 3.1.3. ***(Complete Inner Product Space)***
An inner product space is complete if every Cauchy sequence in the space converges to a limit that is also within the space. In other words, for any Cauchy sequence $\{v_n\}_{n=1}^{\infty}$ *in the inner product space, there exists a vector* v *in the space such that* v_n *converges to* v*.*

Definition 3.1.4. ***(Separable Inner Product Space)***
An inner product space is separable if it contains a countable dense subset. In other words, there exists a countable subset $\{v_1, v_2, v_3, \ldots\}$ *of the space such that any vector in the space can be arbitrarily well-approximated by vectors from this subset.*

Definition 3.1.5. ***(Hilbert Space)***
A Hilbert Space $\mathcal{H}$ *is an inner product space that is complete and separable with respect to the norm defined by the inner product.*

Definition 3.1.6. ***(Kernel Function)***
A kernel function K *is a mapping that takes two input variables,* x *and* y *, and returns a real-valued similarity measure* $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$*, where* $\mathcal{X}$ *represents the input space, which could be finite-dimensional or infinite-dimensional. Key properties of kernel functions include:*

(a) *Symmetry:* $K(x, y) = K(y, x)$

(b) *Positive Semidefinite (p.s.d.):* $\forall x_1, \ldots, x_n \in \mathcal{X}$*, the matrix* $\boldsymbol{K}$ *defined by* $\boldsymbol{K}_{ij} = K(x_i, x_j)$ *is positive semidefinite (A matrix* $Q \in \mathbb{R}^{n \times n}$ *is p.s.d. if* $\forall x \in \mathbb{R}^n$*,* $x^T Q x \geq 0$*)*

Definition 3.1.7. ***(Reproducing Kernel Hilbert Space)***
$K(\cdot, \cdot)$ *is a reproducing kernel of a Hilbert space* $\mathcal{H}$ *if* $\forall f \in \mathcal{H}$*,* $f(x) = \langle f(\cdot), K(\cdot, x) \rangle$

Proposition 3.1.1.
if $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ *is a scalar product, and* $K : V \times V \to \mathbb{R}$ *is a kernel function, then the following holds:*

$$\langle K(\cdot, x), K(\cdot, y) \rangle = K(x, y) \text{ for all } x, y \in V$$

Proof:
It is known that $K(\cdot, x) = \sum_{i=1}^{n} e_i e_i^{\mathrm{T}}(x)$, where $\{e_1, \ldots, e_n\}$ is the canonical orthonormal basis. Then we have:

$$\begin{aligned}
\langle K(\cdot, x), K(\cdot, y) \rangle &= K^{\mathrm{T}}(\cdot, x) K(\cdot, y) = \\
&= \left(\sum_{i=1}^{n} e_i^{\mathrm{T}} e_i(x) \right) \left(\sum_{j=1}^{n} e_j e_j^{\mathrm{T}}(y) \right) = \\
&= \sum_{i,j=1}^{n} e_i(x) e_j^{\mathrm{T}}(y) \underbrace{e_i^{\mathrm{T}} e_j}_{\delta_{ij}} = \\
&= \sum_{i=1}^{n} e_i(x) e_i^{\mathrm{T}}(y) = K(x, y).
\end{aligned}$$

Therefore, the proposition is proven. □

3.2 Positive definite kernels on graphs

We are interested in kernel functions $K : V \times V \to \mathbb{R}$ defined on the graph G, which are symmetric, meaning they satisfy $K(v, w) = K(w, v)$ for all the nodes $v, w \in V$. A kernel K facilitates the introduction of a linear operator $\mathbf{K} : \mathcal{L}(G) \to \mathcal{L}(G)$ which acts on a graph signal $x \in \mathcal{L}(G)$ as follows:

$$\mathbf{K}x(v_i) = \sum_{j=1}^{n} K(v_i, v_j) x(v_j).$$

Given our representation of signals $x \in \mathcal{L}(G)$ as vectors in $\mathbb{R}^n$, we can represent $\mathbf{K}$ as the symmetric matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$, defined by:

$$\mathbf{K} = \begin{pmatrix} K(v_1, v_1) & K(v_1, v_2) & \cdots & K(v_1, v_n) \\ K(v_2, v_1) & K(v_2, v_2) & \cdots & K(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(v_n, v_1) & K(v_n, v_2) & \cdots & K(v_n, v_n) \end{pmatrix}.$$

The subsequent families of symmetric kernels are relevant in our analysis:

Definition 3.2.1.

(a) We call a symmetric kernel K positive semidefinite (p.s.d.) if the matrix $\boldsymbol{K} \in \mathbb{R}^{n \times n}$ is positive semi-definite, i.e., we have $x^T \boldsymbol{K} x \geq 0$ for all $x \in \mathbb{R}^n$.

(b) We call a symmetric kernel K positive definite (p.d.) if the matrix $\boldsymbol{K} \in \mathbb{R}^{n \times n}$ is strictly positive, i.e., we have $x^T \boldsymbol{K} x > 0$ for all $x \in \mathbb{R}^n$, $x \neq 0$.

(c) We call a symmetric kernel K conditionally positive definite (c.p.d.) with respect to a subspace $\mathcal{Y} \subset \mathcal{L}(G)$, if $\boldsymbol{K}$ is p.d. on the subspace $\mathcal{Y}$.

3.3 Interpolation with positive definite kernels

Every positive definite kernel K allows the space $\mathcal{L}(G)$ to be endowed with a scalar product expressed as:

$$\langle x, y \rangle_K = y^{\mathrm{T}} \mathbf{K}^{-1} x, \text{ for all signals } x, y \in \mathcal{L}(G).$$

The resulting inner product space, also referred to as the native space

$$\mathcal{N}_K = \left\{ x \in \mathcal{L}(G) \mid x(v) = \sum_{k=1}^{n} c_k K(v, v_k), v_k \in V, c_k \in \mathbb{R} \right\},$$

constitutes a reproducing kernel Hilbert space [34], wherein K assumes the role of the reproducing kernel satisfying the following property:

$$\langle x, K(\cdot, v_j) \rangle_K = x(v_j) \quad \text{for all signal } x \in \mathcal{L}(G). \tag{3.1}$$

Proof of property (3.1)**:**

$$\langle x, K(\cdot, v_j)\rangle_K = x^{\mathrm{T}}\mathbf{K}^{-1}K(\cdot, v_j) = x^{\mathrm{T}}(\mathbf{K}^{-1})^{\mathrm{T}}K(\cdot, v_j) =$$

$$= \langle \mathbf{K}^{-1}x, K(\cdot, v_j)\rangle = \sum_{k=1}^{n} \underbrace{x(v_k)^{\mathrm{T}}\mathbf{K}^{-1}(\cdot, v_k)}_{a_k} \underbrace{K(v_k, v_j)}_{K(v_j, v_k)} = x(v_j)$$

Thus, the property holds true. □

Proposition 3.3.1.
The scalar product $\langle x, y\rangle_K$ satisfies all the properties in definition (3.1.2)

Proof:
(a) Symmetry

$$\langle x, y\rangle_K = \underbrace{y^{\mathrm{T}}\mathbf{K}^{-1}x}_{\alpha\in\mathbb{R},\ \alpha=\alpha^{\mathrm{T}}} = (y^{\mathrm{T}}\mathbf{K}^{-1}x)^{\mathrm{T}} = x^{\mathrm{T}}\mathbf{K}^{-1}y = \langle y, x\rangle_K.$$

(b) Linearity w.r.t. first term

$$\langle x+y, z\rangle_K = z^{\mathrm{T}}\mathbf{K}^{-1}(x+y) = z^{\mathrm{T}}\mathbf{K}x + z^{\mathrm{T}}\mathbf{K}y = \langle x, z\rangle_K + \langle y, z\rangle_K.$$

(c) Linearity w.r.t. second term

$$\langle x, y+z\rangle_K = (y+z)^{\mathrm{T}}\mathbf{K}^{-1}x = y^{\mathrm{T}}\mathbf{K}x + z^{\mathrm{T}}\mathbf{K}x = \langle x, y\rangle_K + \langle x, z\rangle_K.$$

(d) Associative

$$\langle \alpha x, y\rangle_K = y^{\mathrm{T}}\mathbf{K}^{-1}\alpha x = \alpha y^{\mathrm{T}}\mathbf{K}^{-1}x = \alpha\langle x, y\rangle_K.$$

(e) Positive Definite

$\langle x, x\rangle_K = x^{\mathrm{T}}\mathbf{K}^{-1}x > 0\,\forall x \in \mathcal{L}(G),\ x \neq 0$ by construction: since K is a positive-definite kernel function ⇒ the kernel matrix $\mathbf{K}$ is p.d. ⇒ the inverse kernel matrix $\mathbf{K}^{-1}$ is p.d. too ⇒ $x^{\mathrm{T}}\mathbf{K}^{-1}x > 0\,\forall x \in \mathcal{L}(G),\ x \neq 0.$ □

Proposition 3.3.2. ***(Cauchy-Schwarz inequality)***
For the native space norm $\|\cdot\|_K$, along with the inner product $\langle\cdot,\cdot\rangle_K$, the Cauchy-Schwarz inequality holds true:

$$|\langle x, y\rangle_K| \leq \|x\|_K\|y\|_K$$

Proof:

If one of the two vectors is zero then both sides are zero so we may assume that both x, y are non-zero. Let $t \in \mathbb{R}$. Then:

$$\begin{aligned}
0 \leq &\|x+ty\|_K^2 = \langle x+ty, x+ty\rangle_K = \\
&= \langle x, x\rangle_K + \langle x, ty\rangle_K + \langle ty, x\rangle_K + \langle ty, ty\rangle_K = \\
&= \langle x, x\rangle_K + t\langle x, y\rangle_K + t\underbrace{\langle y, x\rangle_K}_{\langle x,y\rangle_K} + t^2\langle y, y\rangle_K = \\
&= \langle x, x\rangle_K + 2t\langle x, y\rangle_K + t^2\langle y, y\rangle_K = \dots
\end{aligned}$$

Now we choose $t := -\frac{\langle x,y\rangle_K}{\langle y,y\rangle_K}$. Therefore we get:

$$\begin{aligned}
\dots &= \langle x, x\rangle_K - 2\frac{\langle x, y\rangle_K}{\langle y, y\rangle_K}\langle x, y\rangle_K + \frac{\langle x, y\rangle_K^2}{\langle y, y\rangle_K^2}\langle y, y\rangle_K = \\
&= \|x\|_K^2 - \frac{\langle x, y\rangle_K^2}{\langle y, y\rangle_K} = \|x\|_K^2 - \frac{\langle x, y\rangle_K^2}{\|y\|_K^2} \Rightarrow \\
&\Rightarrow 0 \leq \|x\|_K^2 - \frac{\langle x, y\rangle_K^2}{\|y\|_K^2} \Rightarrow \langle x, y\rangle_K^2 \leq \|x\|_K^2\|y\|_K^2 \Rightarrow \\
&\Rightarrow |\langle x, y\rangle_K| \leq \|x\|_K\|y\|_K
\end{aligned}$$

Thus, the Cauchy-Schwarz inequality is proven. □

Proposition 3.3.3.

The native space norm $\|\cdot\|_K$ *satisfies all the properties in definition* (3.1.1)

Proof:

(a) Non-negative

Since K is a positive-definite kernel, the scalar product $\langle x, x\rangle_K \geq 0 \,\forall\, x \in \mathcal{L}(G) \Rightarrow$
$\Rightarrow \|x\|_K = \sqrt{\langle x, x\rangle_K} \geq 0 \,\forall\, x \in \mathcal{L}(G)$

(b) Strictly Positive

Since K is a positive-definite kernel, the scalar product $\langle x, x\rangle_K = 0$ if and only if $x = 0 \Rightarrow$
$\Rightarrow \|x\|_K = \sqrt{\langle x, x\rangle_K} = 0$ if and only if $x = 0$

(c) Homogeneous

$$\|\alpha x\|_K = \sqrt{\langle \alpha x, \alpha x\rangle_K} = \sqrt{\alpha^2\langle x, x\rangle_K} = |\alpha|\sqrt{\langle x, x\rangle_K} = |\alpha|\|x\|_K$$

(d) Triangle Inequality

$$0 \leq \|x+y\|_K^2 = \|x\|_K^2 + 2\langle x, y\rangle_k + \|y\|_K^2 \overset{\text{C–S}}{\leq} \|x\|_K^2 + 2\|x\|_K\|y\|_K + \|y\|_K^2 =$$
$$= (\|x\|_K + \|y\|_K)^2 \Rightarrow \|x+y\|_K^2 \leq (\|x\|_K + \|y\|_K)^2 \Rightarrow$$
$$\Rightarrow \|x+y\|_K \leq \|x\|_K + \|y\|_K. \qquad \square$$

A positive definite kernel K can typically be employed to address interpolation problems within interpolation spaces generated by the columns of the matrix $\mathbf{K}$. The corresponding interpolation problem on the graph G is formulated as follows: given samples $x(w_1), \ldots, x(w_N)$ of a signal x on a subset $W = \{w_1, \ldots, w_N\} \subset V$, where $N \leq n$, the aim is to find an interpolating signal $\mathbf{I}_W x \in \mathcal{L}(G)$ that interpolates x at the nodes in W, that is,

$$\mathbf{I}_W x(w_k) = x(w_k) \quad \text{for all } k \in \{1, \ldots, N\}. \tag{3.2}$$

With a positive definite kernel K this interpolation problem can be addressed as follows: we employ the columns $K(\cdot, w_k)$, where $k \in \{1, \ldots, N\}$, of the matrix $\mathbf{K}$ as the interpolation basis. Subsequently, an interpolating signal $\mathbf{I}_W x$, based on the expansion

$$\mathbf{I}_W x(v) = \sum_{k=1}^{N} c_k K(v, w_k) \tag{3.3}$$

has to satisfy the interpolation condition

$$\underbrace{\begin{pmatrix} K(w_1,w_1) & K(w_1,w_2) & \cdots & K(w_1,w_N) \\ K(w_2,w_1) & K(w_2,w_2) & \cdots & K(w_2,w_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(w_N,w_1) & K(w_N,w_2) & \cdots & K(w_N,w_N) \end{pmatrix}}_{\mathbf{K}_W} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} = \begin{pmatrix} x(w_1) \\ x(w_2) \\ \vdots \\ x(w_N) \end{pmatrix}. \tag{3.4}$$

Since $\mathbf{K}$ is positive definite, also the submatrix $\mathbf{K}_W$ is positive definite according to the inclusion principle ([35], Corollary 4.3.15). Consequently, the linear system (3.4) has a unique solution, and the interpolating signal $\mathbf{I}_W x$ can be uniquely expressed in terms of the basis

$\{K(\cdot, w_1), \ldots, K(\cdot, w_N)\}$. We denote the corresponding interpolation space as:

$$\mathcal{N}_{K,W} = \left\{ x \in \mathcal{L}(G) \;\middle|\; x(v) = \sum_{k=1}^{N} c_k K(v, w_k) \right\}.$$

The following result is standard in reproducing kernel Hilbert spaces and explains why these spaces are widely preferred for signal interpolation and approximation.

Proposition 3.3.4. *([17], Corollary 10.25)*
The interpolant $\mathbf{I}_W x$ of x minimizes the native space norm $\|\cdot\|_K$ over all other possible interpolants of x in $\mathcal{L}(G)$ at the nodes W.

Proof:
if a signal y vanishes on W, the reproducing property of the kernel K leads to the identity:

$$\langle y, \mathbf{I}_W x\rangle_K = \langle y, \sum_{k=1}^{N} c_k K(\cdot, w_k)\rangle_K = \sum_{k=1}^{N} c_k \langle y, K(\cdot, w_k)\rangle_K = \sum_{k=1}^{N} c_k y(w_k) = 0.$$

Therefore, if $z \in \mathcal{L}(G)$ is a second interpolant of x at the nodes W we obtain:

$$\begin{aligned} \|\mathbf{I}_W x\|_K^2 &= \langle \mathbf{I}_W x, \mathbf{I}_W x - z + z\rangle_K = \\ &= \underbrace{\langle \mathbf{I}_W x, \mathbf{I}_W x - z\rangle_K}_{0} + \langle \mathbf{I}_W x, z\rangle_K = \\ &= \langle \mathbf{I}_W x, z\rangle_K \overset{\text{C–S}}{\leq} \|\mathbf{I}_W x\|_K \|\mathbf{z}\|_K. \end{aligned}$$

Let's take a closer look at the equation $\langle \mathbf{I}_W x, \mathbf{I}_W x - z\rangle_K = 0$

$$\begin{aligned} \langle \mathbf{I}_W x, \mathbf{I}_W x - z\rangle &= \langle \sum_{k=1}^{N} c_k K(\cdot, w_k), \sum_{j=1}^{N} K(\cdot, w_j) - z\rangle_K = \\ &= \sum_{k=1}^{N} \sum_{j=1}^{K} \left\{ \langle K(\cdot, w_k), K(\cdot, w_j)\rangle_K - \underbrace{\langle K(\cdot, w_k), z\rangle_K}_{z(w_k)} \right\} = \ldots \end{aligned}$$

We examine the term $\langle K(\cdot, w_k), K(\cdot, w_j)\rangle_K$ and considering that z interpolates x at the same nodes W, we obtain the following:

$$\langle K(\cdot, w_k), K(\cdot, w_j)\rangle_K = \sum_{i=1}^{N} \underbrace{K^{-1}(\cdot, w_i)^{\mathrm{T}} K(w_i, w_j)}_{a_i} K(w_i, w_k) = z(w_k).$$

Now, we can conclude:

$$\ldots = \sum_{k=1}^{N} \sum_{j=1}^{K} \{z(w_k) - z(w_k)\} = 0$$

Therefore, the proof of the proposition is complete. □

3.4 Interpolation with conditionally positive definite kernels

If K is conditionally positive definite (c.p.d.) with respect to a subspace $\mathcal{Y}$, the interpolation on the node set W can be carried out in a similar way, once the challenge related to the non-positive definiteness on the orthogonal complement $\mathcal{Y}^{\perp}$ of $\mathcal{Y}$ is solved. If the dimension M of the complement $\mathcal{Y}^{\perp}$ is small and an orthonormal basis $\{y_1^{\perp}, \ldots, y_M^{\perp}\}$ of $\mathcal{Y}^{\perp}$ is provided, this issue can be straightforward by introducing the augmented kernel:

$$K^{(\delta)}(v, w) = K(v, w) + \delta \left(\sum_{i=1}^{M} y_i^{\perp}(v) y_i^{\perp}(w) \right). \tag{3.5}$$

If the parameter $\delta > |\lambda_{min}(\mathbf{K})| \geq 0$ is larger than the absolute value of the smallest eigenvalue of $\mathbf{K}$, then the augmented kernel $K^{(\delta)}$ is positive definite, allowing us to employ the interpolation procedure outlined in the previous section (3.3). In particular we can find a unique interpolation signal $x \in \mathcal{N}_{K^{(\delta)}, W}$, that solved the interpolation problem in (3.2). The interpolation $\mathbf{I}_W x$ has the expansion

$$\mathbf{I}_W x(v) = \sum_{k=1}^{N} c_k K(v, w_k) + \sum_{i=1}^{M} d_i y_i^{\perp}(v), \quad d_i = \delta \sum_{k=1}^{N} c_k y_k^{\perp}(w_k),$$

where the coefficients c_k are the solution of the system (3.4) with respect to the augmented kernel $K^{(\delta)}$. In particular, the interpolation space $\mathcal{N}_{K^{\delta}, W}$ constitutes an N-dimensional subspace

of the space $\mathcal{N}_{K,W} + \mathcal{Y}^{\perp}$.
Furthermore, the sum $\sum_{i=1}^{M} y_i^{\perp}(v) y_i^{\perp}(w)$ in (3.5) can be interpreted as the reproducing kernel of the orthogonal complement $\mathcal{Y}^{\perp}$. By introducing a δ-multiple of this kernel to the conditionally positive definite kernel K, the non-positive part of the spectrum of K is shifted by $\delta > 0$ in positive direction, resulting in a positive definite kernel $K^{(\delta)}$. An alternative approach to obtaining a positive definite kernel from a conditionally positive definite kernel is outlined in [36], involving a reflection technique and Pontryagin spaces. A third option, as pursued in [13], involves adding a multiple of the identity matrix to the conditionally positive definite kernel K, thus shifting the entire spectrum of the kernel to the positive real axis.
On the other hand, the definition (3.2.1) **(c)** for conditionally positive definite kernels is not the standard definition given in the literature, see for instance [17]. The standard definition reads as follows: K is conditionally positive definite if and only if for all subsets W the matrix $\mathbf{K}_W$ is positive definite on the subspace determined by $\sum_{k=1}^{N} c_k y^{\perp} = 0$ for all $y^{\perp} \in \mathcal{Y}^{\perp}$. As shown in [36], every kernel that is conditionally positive definite with respect to this standard definition can be interpreted as a kernel that is conditionally positive definite with respect to definition (3.2.1) **(c)** and vice versa. Some literature refers to the conditionally positive definite kernels in definition (3.2.1) **(c)** as kernels with a finite number of negative squares, as noted in [22], [36].

3.5 Interpolation with positive definite kernels via Laplacian and Schur complements

Before delving into the details, we recall some theoretical concepts about Schur complements. Let M be an $n \times n$ matrix written as follows

$$\mathrm{M} = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A is a $k \times k$ matrix and D is a $(n-k) \times (n-k)$ matrix, (B is a $k \times (n-k)$ matrix and C is a $(n-k) \times k$ matrix). The linear system can be addressed and solved

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix},$$

that is

$$Ax + By = c,$$
$$Cx + Dy = d,$$

by mimicking Gaussian elimination, namely, assuming that D is invertible, we first solve for y getting

$$y = D^{-1}(d - Cx),$$

and after substituting this expression for y in the first equation, we get

$$Ax + B(D^{-1}(d - Cx)) = c,$$

that is,

$$(A - BD^{-1}C)x = c - BD^{-1}d.$$

If the matrix $A - BD^{-1}C$ is invertible, then we obtain the solution to our system

$$x = (A - BD^{-1}C)^{-1}(c - BD^{-1}d),$$
$$y = D^{-1}\left(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d)\right).$$

The matrix $A - BD^{-1}C$ is referred to as the **Schur Complement** of D in M. If A is invertible, the elimination of x from the first equation reveals that the **Schur Complement** of A in M is $D - CA^{-1}B$.

Definition 3.5.1. ***(Schur Complements)***
Given M, an $n \times n$ matrix decomposed into blocks, with $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ where A and D are square matrices (A is a $k \times k$ matrix and D is an $(n-k) \times (n-k)$ matrix). Then the following holds:

(a) *If A is invertible, the Schur complement of A in M is defined as* ***M/A*** $= D - CA^{-1}B$

(b) *If D is invertible, the Schur complement of D in M is defined as* ***M/D*** $= A - BD^{-1}C$

Proposition 3.5.1.
Given $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ *where A is invertible, then the following holds:*

$$det(M) = det(A)det(M/A)$$

Proof:

$$\begin{pmatrix} \mathrm{I} & 0 \\ -CA^{-1} & \mathrm{I} \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & M/A \end{pmatrix} \implies$$
$$\implies \underbrace{\det\left(\begin{pmatrix} \mathrm{I} & 0 \\ -CA^{-1} & \mathrm{I} \end{pmatrix}\right)}_{=1} \det(\mathrm{M}) = \det(A)\det(M/A) \implies$$
$$\implies \det(\mathrm{M}) = \det(A)\det(M/A)$$

This concludes the proof. □

Proposition 3.5.2.
Given $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ *where D is invertible, then the following holds:*

$$det(M) = det(D)det(M/D)$$

Proof:

$$\begin{pmatrix} \mathrm{I} & -BD^{-1} \\ 0 & \mathrm{I} \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} M/D & 0 \\ C & D \end{pmatrix} \implies$$
$$\implies \underbrace{\det\left(\begin{pmatrix} \mathrm{I} & -BD^{-1} \\ 0 & \mathrm{I} \end{pmatrix}\right)}_{=1} \det(\mathrm{M}) = \det(M/D)\det(D) \implies$$
$$\implies \det(\mathrm{M}) = \det(M/D)\det(D)$$

Then, the proposition is proven. □

Proposition 3.5.3.
A matrix A *is positive definite if and only if* $\exists\, B$ *positive definite such that* $A = B^2$

Proof:
($\Longrightarrow$) A positive definite $\Rightarrow$ A symmetric $\Rightarrow$ A diagonalizable $\Rightarrow$ $\exists\, U, D$ such that $A = UDU^{\mathrm{T}}$, where $D = \mathrm{diag}(d_1, \dots, d_n)$ is a diagonal matrix with $d_i > 0 \,\forall\, i \in \{1, \dots, n\}$ and $UU^{\mathrm{T}} = U^{\mathrm{T}}U = \mathbf{I}_n$ is a orthonormal matrix. Thus

$$A = \underbrace{\left(U\sqrt{D}U^{\mathrm{T}}\right)}_{=B}\left(U\sqrt{D}U^{\mathrm{T}}\right) = B^2.$$

($\Longleftarrow$) $A = B^2$ positive definite $\Rightarrow$ the eigenvalues of A are the squares of the eigenvalues of B $\Rightarrow$ A is positive definite. $\square$

Proposition 3.5.4.
Let A be a $n \times n$ matrix, then the determinant of A is equal to the product of its eigenvalues.

Proof:
Suppose that $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. Then the λ's are also the roots of the characteristic polynomial, i.e.

$$\begin{aligned}\det(A - \lambda\mathbf{I}_n) = p(\lambda) &= (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) = \\ &= (-1)(\lambda - \lambda_1)(-1)(\lambda - \lambda_2)\cdots(-1)(\lambda - \lambda_n) = \\ &= (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda).\end{aligned}$$

The first equality arises from the factorization of a polynomial based on its roots; the leading (highest degree) coefficient $(-1)^n$ can be obtained by expanding the determinant along the diagonal. Now, by setting λ to zero (since it is just a variable), we obtain $\det(A)$ on the left side and $\lambda_1\lambda_2\cdots\lambda_n$ on the right side, that is, we obtain the desired result

$$\det(A) = \lambda_1\lambda_2\cdots\lambda_n.$$

Thus, the determinant of the matrix is equal to the product of its eigenvalues. $\square$

The next step is to present a characterization of symmetric positive definite matrices through Schur complements. Let M be a symmetric matrix, meaning that A, D are symmetric and

$C = B^T$. In this case, M can be expressed as follows:

$$\mathrm{M} = \begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix} = \begin{pmatrix} \mathrm{I} & 0 \\ B^{\mathrm{T}}A^{-1} & \mathrm{I} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - B^{\mathrm{T}}A^{-1}B \end{pmatrix} \begin{pmatrix} \mathrm{I} & 0 \\ B^{\mathrm{T}}A^{-1} & \mathrm{I} \end{pmatrix}^{\mathrm{T}} \tag{3.6}$$

This shows that M is similar to a block-diagonal matrix, with the Schur complement, $C - B^T A^{-1} B$, being symmetric. Consequently, the following version of "Schur's trick" can be applied to verify whether M is positive definite or positive semidefinite. Remember that the respective notations for the positive definite and positive semidefinite matrices are: $\mathrm{M} \succ 0$ and $\mathrm{M} \succeq 0$.

Theorem 3.5.1.
Let M be a $n \times n$ symmetric matrix, where A is a $k \times k$ symmetric matrix and D is a $(n - k) \times (n - k)$ symmetric matrix. Thus, the following statements are equivalent:

(a) *M is positive definite*

(b) *A is positive definite and $D - B^T A^{-1} B$ is positive definite.*

(c) *A and D are positive definite and $\rho(B^T A^{-1} B D^{-1}) < 1$.*

Proof:
[(a)⇔(b)] From (3.6) observe that

$$\begin{pmatrix} \mathrm{I} & 0 \\ B^{\mathrm{T}}A^{-1} & \mathrm{I} \end{pmatrix}^{-1} = \begin{pmatrix} \mathrm{I} & 0 \\ -B^{\mathrm{T}}A^{-1} & \mathrm{I} \end{pmatrix},$$

and for any symmetric matrix, T, and any invertible matrix, P, the matrix T is positive definite ($T \succ 0$) if and only if the matrix PTP^{T}, which is symmetric, is positive definite ($PTP^{\mathrm{T}} \succ 0$). Furthermore, a block diagonal matrix is positive definite if and only if each diagonal block is positive definite. Thus from (3.6), M is positive definite if and only if both A and its Schur complement $D - B^{\mathrm{T}} A^{-1} B$ are positive definite.
For **(b)⇔(c)** the following proposition is required.

Proposition 3.5.5.
Let A and B be two symmetric matrices, with A being positive definite. The condition $A \succ B$ holds if and only if B is positive semidefinite and $\rho(A^{-1}B) < 1$.

Proof of Proposition (3.5.5)

$$A \succ B \Leftrightarrow \mathrm{I} = A^{-1/2}AA^{-1/2} \succ A^{-1/2}BA^{-1/2} \Leftrightarrow 1 > \lambda_{\max}(A^{-1/2}BA^{-1/2})$$

Now let's take a closer look to this matrix $A^{-1/2}BA^{-1/2}$. Let $x \in \mathbb{R}^n$ and $y = A^{-1/2}x$. We observe that $x^{\mathrm{T}}(A^{-1/2})^{\mathrm{T}}BA^{-1/2}x = (A^{-1/2}x)^{\mathrm{T}}BA^{-1/2}x = y^{\mathrm{T}}By \succeq 0$ (B is positive semidefinite) $\Rightarrow$ $(A^{\mathrm{T}})^{-1/2}BA^{-1/2} = A^{-1/2}BA^{-1/2}$ (A is symmetric) is positive semidefinite, then we have

$$\lambda_{\max}(A^{-1/2}BA^{-1/2}) = \lambda_{\max}(A^{-1/2}A^{-1/2}B) = \lambda_{\max}(A^{-1}B).$$

Currently, form Proposition (3.5.3) exists R positive definite such that $A = R^2 = RR^{\mathrm{T}} \Rightarrow R^{-1}A(R^{-1})^{\mathrm{T}} = \mathrm{I}$. Now, the matrix $R^{\mathrm{T}}BR$ is symmetric $\Rightarrow$ $R^{\mathrm{T}}BR$ is diagonalizable $\Rightarrow$ exists an orthonormal matrix U such that $U^{\mathrm{T}}R^{\mathrm{T}}BRU = (RU)^{\mathrm{T}}B(RU) = \Lambda$ is diagonal. Let $S = RU$. Then $S^{-1}A(S^{-1})^{\mathrm{T}} = (RU)^{-1}A((RU)^{-1})^{\mathrm{T}} = U^{-1}R^{-1}A(U^{-1}R^{-1})^{\mathrm{T}} = U^{\mathrm{T}}R^{-1}A(U^{\mathrm{T}}R^{-1})^{\mathrm{T}} = U^{\mathrm{T}}R^{-1}A(R^{-1})^{\mathrm{T}}U = U^{\mathrm{T}}U = \mathrm{I}$ and $SBS^{\mathrm{T}} = (RU)^{\mathrm{T}}B(RU) = U^{\mathrm{T}}R^{\mathrm{T}}BRU = \Lambda$. Thus, $B = (S^{-1})^{\mathrm{T}}\Lambda S^{-1}$ and $A = SS^{\mathrm{T}}$. Furthermore $AB = SS^{\mathrm{T}}(S^{-1})^{\mathrm{T}}$ $\Lambda S^{-1} = S\Lambda S^{-1}$ is diagonalizable. In addition, since B is positive semidefinite, all the eigenvalues of Λ are real and non-negative.
Finally, from this last observation, we can concluded that AB^{-1} has real nonnegative eigenvalues, so $\lambda_{\max}(A^{-1}B) = \rho(A^{-1}B)$. □

[(b)⟹(c)] A is positive definite and $D - B^{\mathrm{T}}A^{-1}B$ is positive definite $\Rightarrow D - B^{\mathrm{T}}A^{-1}B \succ 0 \Rightarrow D \succ B^{\mathrm{T}}A^{-1}B \Rightarrow$ from Proposition (3.5.4), $B^{\mathrm{T}}A^{-1}B$ is positive semidefinite and $\rho(D^{-1}B^{\mathrm{T}}A^{-1}B) = \rho(B^{\mathrm{T}}A^{-1}BD^{-1}) < 1$. In addition, D is positive definite since $B^{\mathrm{T}}A^{-1}B$ is positive semidefinite and $D \succ B^{\mathrm{T}}A^{-1}B$.

[(b)⟸(c)] A, D are positive definite and $\rho(B^{\mathrm{T}}A^{-1}BD^{-1}) < 1 \Rightarrow B^{\mathrm{T}}A^{-1}B$ is positive semidefinite and $D \succ B^{\mathrm{T}}A^{-1}B \Rightarrow$ A is positive definite and $D - B^{\mathrm{T}}A^{-1}B \succ 0 \Rightarrow A$ and $D - B^{\mathrm{T}}A^{-1}B$ are positive definite.
This concludes the proof of Theorem (3.5.1). □

Theorem 3.5.2.
Let M be a $n \times n$ symmetric matrix, where A is a $k \times k$ symmetric matrix and D is a $(n-k) \times (n-k)$ symmetric matrix. Thus, the following statements are equivalent:

- ***(a)*** *M is positive semidefinite*
- ***(b)*** *A is positive semidefinite and $D - B^T A^{-1}B$ is positive semidefinite.*

Proof:
[(a)⇔(b)] From (3.6) observe that

$$\begin{pmatrix} \mathrm{I} & 0 \\ B^{\mathrm{T}}A^{-1} & \mathrm{I} \end{pmatrix}^{-1} = \begin{pmatrix} \mathrm{I} & 0 \\ -B^{\mathrm{T}}A^{-1} & \mathrm{I} \end{pmatrix}$$

and for any symmetric matrix, T, and any invertible matrix, P, the matrix T is positive semidefinite ($T \succeq 0$) if and only if the matrix PTP^{T}, which is symmetric, is positive semidefinite ($PTP^{\mathrm{T}} \succeq 0$). Furthermore, a block diagonal matrix is positive semidefinite if and only if each diagonal block is positive semidefinite. Thus from (3.6), M is positive semidefinite if and only if both A and its Schur complement $D - B^{\mathrm{T}}A^{-1}B$ are positive semidefinite. □

Now, we aim to extend the interpolation method illustrated in section (3.3), taking into account not only the signal values at certain nodes but also the information related to the Laplacian of the signal at each node. Thus, the corresponding problem on the graph G is formulated as follows: given samples $x(w_1), \ldots, x(w_N)$ of a signal x on a subset $W = \{w_1, \ldots, w_N\} \subset V$, where $N \leq n$, the aim is to find the interpolating signals $\mathbf{I}_W x$, $\mathbf{I}_W(\mathbf{L}x) \in \mathcal{L}(G)$ that interpolates x and $\mathbf{L}x$ at the nodes in W in the following way:

$$\begin{cases} \mathbf{I}_W x(w_k) = x(w_k) \quad \text{for all } k \in \{1, \ldots, N\}, \\ \\ \mathbf{I}_W(\mathbf{L}x(w_k)) = \mathbf{L}x(w_k) \quad \text{for all } k \in \{1, \ldots, N\} \end{cases}.$$

With a positive definite kernel K this interpolation problem can be addressed as follows: we employ the columns $K(\cdot, w_k)$, where $k \in \{1, \ldots, N\}$, of the matrix **K** as the interpolation basis. Subsequently, the interpolating signals $\mathbf{I}_W x$ and $\mathbf{I}_W(\mathbf{L}x)$, based on the expansions

$$\begin{cases} \mathbf{I}_W x(v) = \sum_{k=1}^{N} a_k K(v, w_k) + \sum_{k=1}^{N} b_k \mathbf{L}K(v, w_k) \\ \\ \mathbf{I}_W\left(\mathbf{L}x(v)\right) = \sum_{k=1}^{N} a_k \mathbf{L}K(v, w_k) + \sum_{k=1}^{N} b_k \mathbf{L}^2 K(v, w_k) \end{cases}$$

has to satisfy the interpolation condition

$$\underbrace{\begin{pmatrix} K(w_1,w_1) & \cdots & K(w_1,w_N) & \mathbf{L}K(w_1,w_1) & \cdots & \mathbf{L}K(w_1,w_N) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ K(w_N,w_1) & \cdots & K(w_N,w_N) & \mathbf{L}K(w_N,w_1) & \cdots & \mathbf{L}K(w_N,w_N) \\ \mathbf{L}K(w_1,w_1) & \cdots & \mathbf{L}K(w_1,w_N) & \mathbf{L}^2K(w_1,w_1) & \cdots & \mathbf{L}^2K(w_1,w_N) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}K(w_N,w_1) & \cdots & \mathbf{L}K(w_N,w_N) & \mathbf{L}^2K(w_N,w_1) & \cdots & \mathbf{L}^2K(w_N,w_N) \end{pmatrix}}_{\mathbf{\Lambda}} \begin{pmatrix} a_1 \\ \vdots \\ a_N \\ b_1 \\ \vdots \\ b_N \end{pmatrix} = \begin{pmatrix} x(w_1) \\ \vdots \\ x(w_N) \\ \mathbf{L}x(w_1) \\ \vdots \\ \mathbf{L}x(w_N) \end{pmatrix}$$

Observe that $\mathbf{\Lambda}$ is symmetric. Alternatively, the matrix-vector product can be expressed in a more compact form

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{K}_W & \mathbf{L}\mathbf{K}_W \\ \mathbf{L}\mathbf{K}_W & \mathbf{L}^2\mathbf{K}_W \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\big|_W \\ \mathbf{L}\mathbf{X}\big|_W \end{pmatrix}, \tag{3.7}$$

where $\mathbf{a} = (a_1, \ldots, a_N)$, $\mathbf{b} = (b_1, \ldots, b_N)$ and $\mathbf{X}\big|_W = (x(w_1), \ldots, x(w_N))$. As before, since $\mathbf{K}$ is positive definite, also the submatrix $\mathbf{K}_W$ is positive definite according to the inclusion principle ([35], Corollary 4.3.15). However, according to Theorem (3.5.1), the fact that $\mathbf{K}_W$ is positive definite is not sufficient to guarantee the positive definiteness of $\mathbf{\Lambda}$.

Theorem 3.5.3.
A matrix $\mathbf{M}$ is positive semidefinite if and only if the principal submatrix of $\mathbf{M}$ with all maximally linearly independent columns (and corresponding rows) left is positive definite

Proof:
($\Longleftarrow$) Suppose without loss of generality that $\mathbf{M}$ is symmetric and that $\widetilde{\mathbf{M}}$ is the leading $r \times r$ principal submatrix. We have

$$\mathbf{M} = \begin{pmatrix} \widetilde{\mathbf{M}} & B \\ B^{\mathrm{T}} & C \end{pmatrix}.$$

We note that $\mathbf{M}$ is positive semidefinite if both $\widetilde{\mathbf{M}}$ and $\mathbf{M}/\widetilde{\mathbf{M}}$ are positive semidefinite according to Theorem (3.5.2), where $\mathbf{M}/\widetilde{\mathbf{M}}$ is the Schur complement. Therefore,

$$\mathbf{M}/\widetilde{\mathbf{M}} = C - B^{\mathrm{T}}\widetilde{\mathbf{M}}^{-1}B.$$

On the other hand, the rank of $\mathbf{M}$ must be r, so we have

$$r = \text{rank}(\mathbf{M}) = \text{rank}(\widetilde{\mathbf{M}}) + \text{rank}(\mathbf{M}/\widetilde{\mathbf{M}}) = r + \text{rank}(\mathbf{M}/\widetilde{\mathbf{M}}).$$

Thus, $\mathbf{M}/\widetilde{\mathbf{M}} = 0$, which means that $\mathbf{M}/\widetilde{\mathbf{M}}$ is positive semidefinite. Hence, we conclude that $\mathbf{M}$ is indeed positive semidefinite.

($\Longrightarrow$) Suppose without loss of generality that the first r columns of $\mathbf{M}$ form the maximal linearly independent set in question. That is, $\mathbf{M}v \neq 0$ for all v in the span of $\{e_1, \ldots, e_r\}$ (where e_j is the j-th canonical basis vector). Thus, it follows that $v^{\mathrm{T}}\mathbf{M}v \neq 0$ for all v in the span $\{e_1, \ldots, e_r\}$. Hence, we conclude that the leading $r \times r$ submatrix $\widetilde{\mathbf{M}}$ is positive definite, as required. □

From this last theorem, we can infer that the matrix $\mathbf{\Lambda}$ is positive semidefinite.

Lemma 3.5.1.
The matrices $\boldsymbol{LK}_W$ and $\boldsymbol{L}^2\boldsymbol{K}_W$ are positive semifedinite

Proof:
To prove this lemma, we recall the following:

(a)

$$\|A\|_2 = \sup_{x \neq \bar{0}} \frac{\|Ax\|_2}{\|x\|_2} = \sup_{\|x\|_2 = 1} \|Ax\|_2$$

(b)

$$\lambda_{\text{min}}(A) = \min_{x \neq \bar{0}} \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}x}, \text{ and } \lambda_{\text{max}}(A) = \max_{x \neq \bar{0}} \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}x}$$

In particular, the eigenvalues of $\mathbf{L}$ lie between $[0, 2]$ which implies that the eigenvalues of $\mathbf{L}^2$ lie between $[0, 4]$. Moreover, we know that all the eigenvalues of the submatrix $\mathbf{K}_W$ are positive, as K is a positive definite kernel. We know that $\lambda_{\text{min}}(A)\|x\|_2 \leq \|Ax\|_2 \leq \lambda_{\text{max}}(A)\|x\|_2$. Taking the supremum ($\|x\|_2 = 1$) yields $\lambda_{\text{min}}(A) \leq \|A\|_2 \leq \lambda_{\text{max}}(A)$. In our scenario, we have $\lambda_{\text{min}}(\mathbf{LK}_W) \leq \|\mathbf{LK}_W\|_2 \leq \lambda_{\text{max}}(\mathbf{LK}_W)$. Now, we proceed to show that $\lambda_{\text{min}}(AB) \geq \lambda_{\text{min}}(A)\lambda_{\text{min}}(B)$ and $\lambda_{\text{max}}(AB) \leq \lambda_{\text{max}}(A)\lambda_{\text{max}}(B)$, where A is a semidefinite positive matrix and B is a positive definite matrix. Since B is positive definite, according to Proposition (3.5.3), there exists a symmetric positive definite matrix C such that $B = C^2$. Thus

$$\det(AB - \lambda \mathbf{I}_n) = \det(AC^2 - \lambda \mathbf{I}_n) = \det\left(C(AC^2 - \lambda \mathbf{I}_n)C^{-1}\right) = \\ = \det(CAC - \lambda C \mathbf{I}_n C^{-1}) = \det(CAC - \lambda \mathbf{I}_n).$$

It follows that the characteristic polynomial of AB is the same as that of CAC. Therefore, the minimum and maximum eigenvalues of AB match those of CAC, which are $\min_{x\neq\bar{0}} \frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}x}$ and $\max_{x\neq\bar{0}} \frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}x}$. Now, $\forall\, x \neq \bar{0}$, we have:

$$\frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}x} = \frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}CCx} \cdot \frac{x^{\mathrm{T}}CCx}{x^{\mathrm{T}}x} \geq \\ \geq \left(\min_{y\neq\bar{0}} \frac{y^{\mathrm{T}}Ay}{y^{\mathrm{T}}y}\right)\left(\min_{x\neq\bar{0}} \frac{x^{\mathrm{T}}Bx}{x^{\mathrm{T}}x}\right) = \lambda_{\text{min}}(A)\lambda_{\text{min}}(B),$$

$$\frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}x} = \frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}CCx} \cdot \frac{x^{\mathrm{T}}CCx}{x^{\mathrm{T}}x} \leq \\ \leq \left(\max_{y\neq\bar{0}} \frac{y^{\mathrm{T}}Ay}{y^{\mathrm{T}}y}\right)\left(\max_{x\neq\bar{0}} \frac{x^{\mathrm{T}}Bx}{x^{\mathrm{T}}x}\right) = \lambda_{\text{max}}(A)\lambda_{\text{max}}(B),$$

with $y = Cx$, $y^{\mathrm{T}} = (Cy)^{\mathrm{T}} = x^{\mathrm{T}}C$. Then

$$\lambda_{\text{min}}(AB) = \lambda_{\text{min}}(CAC) = \min_{x\neq\bar{0}} \frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}x} \geq \\ \geq \left(\min_{y\neq\bar{0}} \frac{y^{\mathrm{T}}Ay}{y^{\mathrm{T}}y}\right)\left(\min_{x\neq\bar{0}} \frac{x^{\mathrm{T}}Bx}{x^{\mathrm{T}}x}\right) = \lambda_{\text{min}}(A)\lambda_{\text{min}}(B),$$

$$\lambda_{\text{max}}(AB) = \lambda_{\text{max}}(CAC) = \max_{x\neq\bar{0}} \frac{x^{\mathrm{T}}CACx}{x^{\mathrm{T}}x} \leq \\ \leq \left(\max_{y\neq\bar{0}} \frac{y^{\mathrm{T}}Ay}{y^{\mathrm{T}}y}\right)\left(\max_{x\neq\bar{0}} \frac{x^{\mathrm{T}}Bx}{x^{\mathrm{T}}x}\right) = \lambda_{\text{max}}(A)\lambda_{\text{max}}(B).$$

Finally, we have this chain of inequalities

$$0 = \underbrace{\lambda_{\text{min}}(\mathbf{L})}_{=0} \lambda_{\text{min}}(\mathbf{K}_W) \leq \lambda_{\text{min}}(\mathbf{L}\mathbf{K}_W) \leq \|\mathbf{L}\mathbf{K}_W\|_2 \leq \\ \leq \lambda_{\text{max}}(\mathbf{L}\mathbf{K}_W) \leq \underbrace{\lambda_{\textbf{max}}(\mathbf{L})}_{=1} \lambda_{\text{max}}(\mathbf{K}_W) = \lambda_{\text{max}}(\mathbf{K}_W).$$

Since all steps apply equally to $\mathbf{L}^2\mathbf{K}_W$, we can conclude that both matrices, $\mathbf{L}\mathbf{K}_W$ and $\mathbf{L}^2\mathbf{K}_W$, are positive semidefinite. □

Now, form point **(b)** of Theorem (3.5.1), $\mathbf{\Lambda}$ is positive definite if and only if $\mathbf{K}_W$ and its Schur complement $\mathbf{\Lambda}/\mathbf{K}_W = \mathbf{L}^2\mathbf{K}_W - (\mathbf{L}\mathbf{K}_W)\mathbf{K}_W^{-1}(\mathbf{L}\mathbf{K}_W)$ are positive definite. Since $\mathbf{K}_W$ is already positive definite, we focus our attention on the Schur complement of $\mathbf{K}_W$ in $\mathbf{\Lambda}$. If $N < n$, we are unable to reach any conclusions. In particular, if $N = n$, meaning $W = \{w_1, \ldots, w_n\}$ represents the set of all nodes, the associative property of matrices holds, thus leading to

$$\mathbf{\Lambda}/\mathbf{K}_W = \mathbf{L}^2\mathbf{K}_W - \mathbf{L}\underbrace{\mathbf{K}_W\mathbf{K}_W^{-1}}_{\mathbf{I}_n}\mathbf{L}\mathbf{K}_W = \mathbf{L}^2\mathbf{K}_W - \mathbf{L}^2\mathbf{K}_W = 0_n.$$

Consequently, the Schur complement of $\mathbf{K}_W$ in $\mathbf{\Lambda}$ is not positive definite. Nevertheless, we can address the problem in the following way

$$\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda} + \begin{pmatrix} 0_N & 0_N \\ 0_N & \varepsilon\mathbf{I}_N \end{pmatrix} = \begin{pmatrix} \mathbf{K}_W & \mathbf{L}\mathbf{K}_W \\ \mathbf{L}\mathbf{K}_W & (\mathbf{L}^2\mathbf{K}_W)_\varepsilon \end{pmatrix}, \tag{3.8}$$

where ε is a positive constant and $(\mathbf{L}^2\mathbf{K}_W)_\varepsilon = \mathbf{L}^2\mathbf{K}_W + \varepsilon\mathbf{I}_N$. Therefore the linear system in (3.7) becomes

$$\underbrace{\begin{pmatrix} \mathbf{K}_W & \mathbf{L}\mathbf{K}_W \\ \mathbf{L}\mathbf{K}_W & (\mathbf{L}^2\mathbf{K}_W)_\varepsilon \end{pmatrix}}_{\widetilde{\mathbf{\Lambda}}} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\big|_W \\ \mathbf{L}\mathbf{X}\big|_W \end{pmatrix}. \tag{3.9}$$

Then, the Schur complement of $\mathbf{K}_W$ in $\widetilde{\mathbf{\Lambda}}$ is now positive definite

$$\widetilde{\mathbf{\Lambda}}/\mathbf{K}_W = \mathbf{L}^2\mathbf{K}_W + \varepsilon\mathbf{I}_N - \mathbf{L}^2\mathbf{K}_W = \varepsilon\mathbf{I}_N \;\Rightarrow\; \widetilde{\mathbf{\Lambda}}/\mathbf{K}_W \succ 0.$$

Thus, we can conclude that $\widetilde{\mathbf{\Lambda}}$ is positive definite, as stated in point **(b)** of Theorem (3.5.1).

On the other hand, in the scenario where $N < n$, it is more convenient to use point **(c)** of Theorem (3.5.1). We already know that the matrix $\mathbf{L}^2\mathbf{K}_W$ is positive semidefinite: therefore, the transformation obtained in (3.8) ensures the positive definiteness of $(\mathbf{L}^2\mathbf{K}_W)_\varepsilon$. In this scenario as well, we consider the transformation of $\mathbf{\Lambda}$, i.e. $\widetilde{\mathbf{\Lambda}}$. It remains to be determined for which values of ε the $\rho\left[(\mathbf{L}\mathbf{K}_W)\mathbf{K}_W^{-1}(\mathbf{L}\mathbf{K}_W)(\mathbf{L}^2\mathbf{K}_W)_\varepsilon^{-1}\right] < 1$ $(*)$. To find the values of ε, we recall the definition of the spectral radius and the subsequent lemma along with its proof.

Definition 3.5.2. *(Spectral Radius)*
Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of a matrix A. The spectral radius of A is defined as

$$\rho(A) = \max\{|\lambda_1|, \ldots, |\lambda_n|\}.$$

For a positive definite matrix A, it follows that $\rho(A) = \lambda_{\text{max}}(A)$.

Lemma 3.5.2.
Let A be an invertible matrix. If λ is an eigenvalue of the matrix A, then $1/\lambda$ is an eigenvalue of A^{-1}.

Proof:
Since A is invertible $\Rightarrow$ $\det(A) \neq 0 \Rightarrow$ from Proposition (3.5.4) we conclude that all the eigenvalues are non-zero. Then, suppose λ is an eigenvalue of A, therefore $Ax = \lambda x$ for some $x \neq \bar{0}$. Thus

$$Ax = \lambda x \Rightarrow A^{-1}Ax = \lambda A^{-1}x \Rightarrow x = \lambda A^{-1}x \Rightarrow \frac{1}{\lambda}x = A^{-1}x.$$

We can conclude that $1/\lambda$ is an eigenvalue of A^{-1}. □

From this lemma, we can infer that for every eigenvalue λ_i of A, the corresponding eigenvalue of A^{-1} is $1/\lambda_i$. In particular:

- The maximum eigenvalue of A^{-1} is given by:

$$\lambda_{\text{max}}(A^{-1}) = \frac{1}{\lambda_{\text{min}}(A)}$$

- The minimum eigenvalue of A^{-1} is given by:

$$\lambda_{\text{min}}(A^{-1}) = \frac{1}{\lambda_{\text{max}}(A)}$$

At this point, we establish for which values of ε the inequality $(*)$ holds.

$$\begin{aligned}
&\rho\left[(\mathbf{LK}_W)\mathbf{K}_W^{-1}(\mathbf{LK}_W)(\mathbf{L}^2\mathbf{K}_W)_\varepsilon^{-1}\right] = \\
&\quad = \lambda_{\text{max}}\left[(\mathbf{LK}_W)\mathbf{K}_W^{-1}(\mathbf{LK}_W)(\mathbf{L}^2\mathbf{K}_W)_\varepsilon^{-1}\right] \leq \\
&\quad \leq \lambda_{\text{max}}(\mathbf{LK}_W)\lambda_{\text{max}}(\mathbf{K}_W^{-1})\lambda_{\text{max}}(\mathbf{LK}_W)\lambda_{\text{max}}\left[(\mathbf{L}^2\mathbf{K}_W)_\varepsilon^{-1}\right] =
\end{aligned}$$

$$
\begin{aligned}
&= \lambda_{\max}(\mathbf{L}\mathbf{K}_W)\frac{1}{\lambda_{\min}(\mathbf{K}_W)}\lambda_{\max}(\mathbf{L}\mathbf{K}_W)\frac{1}{\lambda_{\min}\left[(\mathbf{L}^2\mathbf{K}_W)_\varepsilon\right]} \leq \\
&\leq \frac{\lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)}{\lambda_{\min}(\mathbf{K}_W)}\frac{1}{\lambda_{\min}(\mathbf{L}^2\mathbf{K}_W + \varepsilon\mathbf{I}_N)} \leq \\
&\leq \frac{\lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)}{\lambda_{\min}(\mathbf{K}_W)[\lambda_{\min}(\mathbf{L}^2\mathbf{K}_W) + \varepsilon\underbrace{\lambda_{\min}(\mathbf{I}_n)}_{1}]} = \\
&= \frac{\lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)}{\lambda_{\min}(\mathbf{K}_W)\left[\lambda_{\min}(\mathbf{L}^2\mathbf{K}_W) + \varepsilon\right]} \Rightarrow \\
&\Rightarrow \rho\left[(\mathbf{L}\mathbf{K}_W)\mathbf{K}_W^{-1}(\mathbf{L}\mathbf{K}_W)(\mathbf{L}^2\mathbf{K}_W)_\varepsilon^{-1}\right] \leq \frac{\lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)}{\lambda_{\min}(\mathbf{K}_W)\left[\lambda_{\min}(\mathbf{L}^2\mathbf{K}_W) + \varepsilon\right]}.
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
\rho\left[(\mathbf{L}\mathbf{K}_W)\mathbf{K}_W^{-1}(\mathbf{L}\mathbf{K}_W)(\mathbf{L}^2\mathbf{K}_W)_\varepsilon^{-1}\right] < 1 \iff & \frac{\lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)}{\lambda_{\min}(\mathbf{K}_W)\left[\lambda_{\min}(\mathbf{L}^2\mathbf{K}_W) + \varepsilon\right]} < 1 \\
\iff & \varepsilon > \frac{\lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)}{\lambda_{\min}(\mathbf{K}_W)} - \lambda_{\min}(\mathbf{L}^2\mathbf{K}_W).
\end{aligned}
$$

Then, in the case where $N < n$ and $\varepsilon > \lambda^2_{\max}(\mathbf{L}\mathbf{K}_W)/\lambda_{\min}(\mathbf{K}_W) - \lambda_{\min}(\mathbf{L}^2\mathbf{K}_W)$, we can conclude that $\widetilde{\mathbf{\Lambda}}$ is positive definite, as mentioned in point **(c)** of Theorem (3.5.1).
Consequently, since $\widetilde{\mathbf{\Lambda}}$ is invertible, the linear system (3.9) has a unique solution for both scenarios just discussed, and the interpolating signals $\mathbf{I}_W x$ and $\mathbf{I}_W(\mathbf{L}x)$ can be uniquely expressed in terms of the basis $\{K(\cdot, w_1), \dots K(\cdot, w_N)\}$. We denote the corresponding interpolant space as:

$$
\mathcal{N}^{\mathbf{L}}_{K,W} = \left\{ x, \mathbf{L}x \in \mathcal{L}(G) \;\middle|\; \begin{array}{l} \text{i) } x(v) = \displaystyle\sum_{k=1}^{N} a_k K(v, w_k) + \sum_{k=1}^{N} b_k \mathbf{L}K(v, w_k) \\ \text{ii) } \mathbf{L}x(v) = \displaystyle\sum_{k=1}^{N} a_k \mathbf{L}K(v, w_k) + \sum_{k=1}^{N} b_k \left[\mathbf{L}^2 K(v, w_k) + \varepsilon \mathbb{1}_{\{v=w_k\}}\right] \end{array} \right\},
$$

where the $\mathbb{1}_{\{v=w_k\}}$ denotes the indicator function, which equals 1 when $v = w_k$ and 0 otherwise.

3.6 Interpolation with positive definite kernels via Laplacian and under ε-separated nodes

Before delving into the details, we recall the definitions of ε-separated nodes and the distance between two nodes on a graph.

Definition 3.6.1. ***(ε-separated nodes)***
Let $\{v_i\}_{i=1}^N$ be a set of nodes in a graph. These nodes are said to be ε-separated if the distance between any two distinct nodes v_i and v_j in the graph is greater than or equal to ε, i.e.,

$$d(v_i, v_j) \geq \varepsilon \quad \textit{for all} \quad i \neq j,$$

where $d(v_i, v_j)$ represents the square norm $\|v_j - v_i\|_2 \geq \varepsilon$ for all $i, j = 1, \ldots, N$, and ε is a positive constant that defines the minimum separation between the nodes.

Definition 3.6.2. ***(Distance between two nodes on graph)***
Let $G = (V, E)$ be a graph, where V is the set of nodes and E is the set of edges. The distance between two nodes $v_i, v_j \in V$, denoted as dist(v_i, v_j), is defined as:

$$\textit{dist}(v_i, v_j) = \min \{|P| \mid P \textit{ is a path from } v_i \textit{ to } v_j\},$$

where:

- *A **path** P is a sequence of vertices $v_0, v_1, \ldots, v_k$ such that for each consecutive pair of vertices (v_{i-1}, v_i), there exists an edge in E connecting them.*
- *$|P|$ is the number of edges in the path P, i.e., the length of the path.*

*In the case of a **weighted graph** with edge weights $w(e) \in \mathbb{R}$ for each $e \in E$, the distance is instead defined as the minimum sum of the weights along any path from v_i to v_j:*

$$\textit{dist}(v_i, v_j) = \min \left\{ \sum_{e \in P} v(e) \mid P \textit{ is a path from } v_i \textit{ to } v_j \right\}.$$

If there is no path between v_i and v_j, the distance is considered infinite:

$$\textit{dist}(v_i, v_j) = \infty.$$

*This definition applies to both **directed** and **undirected** graphs, with the restriction that in directed graphs, the direction of edges must be respected when forming paths.*

Now, the purpose is to extend the interpolation method illustrated in section (3.3), taking into account not only the signal values at certain nodes but also the information related to the Laplacian of the signal at each node. Thus, the interpolation problem is the same as the one defined in the section (3.5)

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{K}_W & \mathbf{LK}_W \\ \mathbf{LK}_W & \mathbf{L}^2\mathbf{K}_W \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\big|_W \\ \mathbf{LX}\big|_W \end{pmatrix},$$

where here $\mathbf{L}$ is the standard Laplacian, $\mathbf{a} = (a_1, \ldots, a_N)$, $\mathbf{b} = (b_1, \ldots, b_N)$ and $\mathbf{X}\big|_W = (x(w_1), \ldots, x(w_N))$. As before, the fact that $\mathbf{K}_W$ is positive definite is not sufficient to guarantee the positive definiteness of $\mathbf{\Lambda}$.

Proposition 3.6.1.
If the distance between two nodes v_i and v_j in a graph with unitary weights (i.e., $\boldsymbol{A}_{ij} = 1$ for all $i, j = 1, \ldots, n$) is 3, then their neighbor sets are disjoint, i.e.,

$$N(v_i) \cap N(v_j) = \emptyset,$$

where $N(v_i)$ denotes the set of direct neighbors of v_i, i.e., those at distance 1.

Proof:
If a node v_j is a direct neighbor of v_i, that is, $v_j \in N(v_i)$, then the distance between v_i and v_j is 1.
If a node v_k is a neighbor of a neighbor of v_i, that is, $v_k \in N(v_j)$ for some $v_j \in N(v_i)$, then the distance between v_i and v_k is 2, passing through the intermediate node v_j.
When the distance between two nodes v_i and v_j is 3, it implies that there is no path between them shorter than 3 edges. Additionally, the set of direct neighbors of v_i, denoted $N(v_i)$, cannot overlap with the set of direct neighbors of v_j, denoted $N(v_j)$, because the nodes at distance 1 from v_j (which include its direct neighbors) cannot include v_i or its direct neighbors. Since each node in $N(v_i)$ is at distance 1 from v_i, while each node in $N(v_j)$ is at least at distance 2 from v_i, it follows that:

$$N(v_i) \cap N(v_j) = \emptyset.$$

Therefore, no node can belong to both $N(v_i)$ and $N(v_j)$, as the distances between the nodes in these two sets are incompatible with the graph's distance definition. In conclusion, a distance of 3 between two nodes in a graph with normalized weights is sufficient to ensure that their direct neighbor sets are disjoint. □

Theorem 3.6.1.
Let $W = \{w_1, \ldots, w_N\} \subset V$ be a set of nodes in the graph G with unitary weights on the edges (i.e., $\boldsymbol{A}_{ij} = 1$ for all $i, j = 1, \ldots, n$). If the nodes in W are 3-separated, then the sets $\{K(\cdot, w_i)\}_{i=1}^N$ and $\{\mathbf{L}K(\cdot, w_i)\}_{i=1}^N$ are linearly independent.

Proof:
The aim is to show that the sets of functions $\{K(\cdot, w_i)\}_{i=1}^N$ and $\{\mathbf{L}K(\cdot, w_i)\}_{i=1}^N$ are linearly independent. To establish this, we must show that the scalars $\{\alpha_i\}_{i=1}^N$ and $\{\beta_i\}_{i=1}^N$ satisfy the equation

$$\sum_{i=1}^{N} \alpha_i K(\cdot, w_i) + \sum_{i=1}^{N} \beta_i (\mathbf{L}K)(\cdot, w_i) = 0 \tag{3.10}$$

if and only if $\alpha_i = \beta_i = 0$ for all $i = 1, \ldots, N$.
In addition, we can leverage the property of the Laplacian $\mathbf{L}$, which allows the set $\{\mathbf{L}K(\cdot, w_i)\}_{i=1}^N$ to be expressed as a linear combination of the kernel functions associated with the neighboring nodes of w_i. Specifically, we have

$$(K\mathbf{L})(\cdot, w_i) = (\mathbf{L}K)(\cdot, w_i) = \sum_{v_j \sim w_i} (K(\cdot, w_i) - K(\cdot, v_j)),$$

where $v_j \sim w_i$ denotes the neighboring nodes of w_i.
Now, by replacing the expression for $(\mathbf{L}K)(\cdot, w_i)$ in (3.10), we obtain

$$\sum_{i=1}^{N} \alpha_i K(\cdot, w_i) + \sum_{i=1}^{N} \beta_i \sum_{v_j \sim w_i} (K(\cdot, w_i) - K(\cdot, v_j)) = 0. \tag{3.11}$$

Since K is a positive definite kernel and $W = \{w_1, \ldots, w_N\} \subset V$, the sets $\{K(\cdot, w_i)\}_{i=1}^N$ and $\{K(\cdot, v_j)\}_{j=1}^n$ are linearly independent, so the equation (3.11) cannot vanish unless all the coefficients $\{\alpha_i\}_{i=1}^N$ and $\{\beta_i\}_{i=1}^N$ are zero. Hence, the sets $\{K(\cdot, w_i)\}_{i=1}^N$ and $\{\mathbf{L}K(\cdot, w_i)\}_{i=1}^N$ are linearly independent. □

Note: The assumption that the nodes are 3-separated plays a crucial role. Indeed, for every $i = 1, \ldots, N$, we have

$$\begin{aligned}\|\mathbf{L}K(\cdot, w_i) - K(\cdot, w_i)\|_2 &\overset{\text{trg. ineq.}}{\leq} \|\mathbf{L}K(\cdot, w_i)\|_2 + \|K(\cdot, w_i)\|_2 \leq \\ &\leq \underbrace{\|\mathbf{L}\|_2}_{=2} \|K(\cdot, w_i)\|_2 + \|K(\cdot, w_i)\|_2 = 3\|K(\cdot, w_i)\|_2.\end{aligned}$$

Thus, this inequality ensures that the Laplacian transformation does not drastically change the "spread" of the kernel functions $K(\cdot, w_i)$. In particular, the transformed kernel function is at most 3 times larger than the original one in terms of its "spread". Hence, if the nodes $\{w_i\}_{i=1}^N$ are separated by a factor of 3 is sufficient to guarantee that the kernel functions $K(\cdot, w_i)$ and $\mathbf{L}K(\cdot, w_i)$ remain linearly independent for all $i = 1, \ldots, N$.

Consequently, by Theorem (3.6.1) the columns (or rows) of the matrix $\mathbf{\Lambda}$ are linearly independent, implying that $\mathbf{\Lambda}$ is invertible. Therefore, the linear system (3.7) has a unique solution, and the interpolating signals $\mathbf{I}_W x$ and $\mathbf{I}_W(\mathbf{L}x)$ can be uniquely expressed in terms of the basis $\{K(\cdot, w_1), \ldots, K(\cdot, w_N)\}$. The corresponding interpolant space is denoted as:

$$\mathcal{N}_{K,W}^{\mathbf{L},3} = \left\{ x, \mathbf{L}x \in \mathcal{L}(G) \;\middle|\; \begin{array}{l} \text{i) } x(v) = \sum_{k=1}^{N} a_k K(v, w_k) + \sum_{k=1}^{N} b_k \mathbf{L}K(v, w_k) \\ \text{ii) } \mathbf{L}x(v) = \sum_{k=1}^{N} a_k \mathbf{L}K(v, w_k) + \sum_{k=1}^{N} b_k \mathbf{L}^2 K(v, w_k) \end{array} \right\},$$

where the factor 3 denotes the separation distance between the nodes in $W = \{w_1, \ldots, w_N\}$.

4

Positive Definite Functions On Graphs

The interpolation scheme outlined in the previous chapter, based on general kernel methods, lacks the incorporation of spectral information from the graph. Thus, the aim of this chapter is to harmonize these two structures and develop an interpolation scheme where the kernels used for interpolation are defined in relation to the generalized translates of a single graph basis function. To achieve this goal, we rely on the crucial concept of positive definiteness.

4.1 Defining the framework

Definition 4.1.1. ***(Positive semidefinite function)***
We call a function $f : V \to \mathbb{R}$ on the graph G positive semidefinite (positive definite) if the matrix

$$\boldsymbol{K}_f = \begin{pmatrix} \boldsymbol{C}_{e_1} f(v_1) & \boldsymbol{C}_{e_2} f(v_1) & \cdots & \boldsymbol{C}_{e_n} f(v_1) \\ \boldsymbol{C}_{e_1} f(v_2) & \boldsymbol{C}_{e_2} f(v_2) & \cdots & \boldsymbol{C}_{e_n} f(v_2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_{e_1} f(v_n) & \boldsymbol{C}_{e_2} f(v_n) & \cdots & \boldsymbol{C}_{e_n} f(v_n) \end{pmatrix}$$

is symmetric and positive semidefinite (positive definite, respectively). We call f conditionally positive definite with respect to a subspace $\mathcal{Y}$ if $\boldsymbol{K}_f$ is p.d. on $\mathcal{Y}$. The sets of positive semidefinite and positive definite functions in $\mathcal{L}(G)$ are denoted by $\mathcal{P}$ and $\mathcal{P}_+$, respectively.

In particular, a positive semidefinite function f induces naturally a p.s.d. kernel $\mathbf{K}_f$ on G by:

$$K_f(v_i, v_j) := \mathbf{C}_{e_j} f(v_i), \text{ with } \mathbf{C}_{e_j} = \mathbf{U}\mathbf{M}_{\hat{e}_j}\mathbf{U}^{\mathrm{T}}.$$

If G has an additional group structure, the signal $\mathbf{C}_{e_i} f$ represents the shift of the signal f by the group element v_i. However, on general graphs, there isn't an intrinsic concept of translation. Nevertheless, we will encounter several examples in which the signals $\mathbf{C}_{e_i} f$ are spatially well-localized around the nodes v_i. Hence, we can interpret $\mathbf{C}_{e_i} f$ as a generalized translate of f on G. Regardless, the positive definiteness of the function f implies that the collection $\{\mathbf{C}_{e_1} f, \ldots, \mathbf{C}_{e_n} f\}$ of generalized shifts of f is linearly independent and constitutes a basis for $\mathcal{L}(G)$.

4.2 A Bochner-type characterization of positive definite functions

The concept of positive definiteness is closely related to the spectrum $\hat{G}$ of the graph G, which consists of the set of eigenvectors $\{u_1, \ldots, u_n\}$ of the graph Laplacian. This relationship is evident in the Bochner-type characterization of a positive definite function f, which can be described in terms of the graph Fourier transform $\hat{f}$.

Theorem 4.2.1. ***(Bochner-type characterization)***
A function $f \in \mathcal{L}(G)$ is contained in $\mathcal{P}$ ($\mathcal{P}_+$) if and only if $\hat{f}_k \geq 0$ ($\hat{f}_k > 0$, respectively) for all $k \in \{1, \ldots, n\}$. The corresponding p.s.d. kernel K_f has the Mercer decomposition

$$K_f(v, w) = \sum_{k=1}^{n} \hat{f}_k u_k(v) u_k(w).$$

Furthermore, we have the following refinements:

(1) $f \in \mathcal{A}_L \cap \mathcal{P}$ if and only if $\hat{f}_k \geq 0$ and $\hat{f}_k = \hat{f}_{k'}$ for $\lambda_k = \lambda_{k'}$.

(2) $f \in \mathcal{B}_M$ if and only if $\hat{f}_k = 0$ for all indices $k > M$.

(3) f is conditionally positive definite with respect to the subspace $\mathcal{Y} =$ span $\{u_{k_1}, \ldots, u_{k_K}\}$ if and only if $\hat{f}_{k_1} > 0, \ldots, \hat{f}_{k_K} > 0$.

Proof:

We apply the definition of the convolution given in (2.7) to the kernel matrix $\mathbf{K}_f$. By applying the convolution operator $\mathbf{C}_x = \mathbf{U}\mathbf{M}_{\hat{x}}\mathbf{U}^{\mathrm{T}}$ and taking advantage of the convolution's commutativity, we can rewrite the columns of the kernel matrix $\mathbf{K}_f$ as follows:

$$\begin{aligned} K_f(\cdot, v_i) = \mathbf{C}_{e_i} f = \mathbf{U}\mathbf{M}_{\hat{e}_i}\mathbf{U}^{\mathrm{T}} f = \mathbf{U}\mathbf{M}_{\hat{e}_i}\hat{f} = \\ = \mathbf{U}\mathbf{M}_{\hat{f}}\,\hat{e}_i = \mathbf{U}\mathbf{M}_{\hat{f}}\mathbf{U}^{\mathrm{T}} e_i = \mathbf{C}_f e_i. \end{aligned}$$

This leads to $\mathbf{K}_f = \mathbf{U}\mathbf{M}_{\hat{f}}\mathbf{U}^{\mathrm{T}}$, revealing the spectral decomposition of the matrix $\mathbf{K}_f$ as well as the Mercer decomposition of $\mathbf{K}_f$:

$$\begin{aligned} K_f(v_i, v_j) = & \mathbf{C}_{e_j} f(v_i) = \mathbf{C}_f \underbrace{e_j(v_i)}_{\delta_{ij}} = \\ & = \mathbf{C}_f = \mathbf{U}\mathbf{M}_{\hat{f}}\mathbf{U}^{\mathrm{T}} = \sum_{k=1}^{n} \hat{f}_k u_k(v_i) u_k(v_j) \end{aligned}$$

Since the entries $\hat{f}_k$ of $\hat{f}$ correspond to the eigenvalues of $\mathbf{K}_f$, we can conclude that $\mathbf{K}_f$ is positive semidefinite (p.d.) if and only if $\hat{f}_k \geq 0$ ($\hat{f}_k > 0$) for all $k \in \{1, \ldots, n\}$. The additional refinements for the subalgebras $\mathcal{A}_{\mathrm{L}}$ and $\mathcal{B}_M$ as well as for conditionally positive definite functions can be directly inferred from its definitions and proposition (2.4.1):

(1) $f \in \mathcal{A}_{\mathrm{L}} \cap \mathcal{P} \iff f$ is positive semidefinite and $\hat{f}_k = \hat{f}_{k'}$ if $\lambda_k = \lambda_{k'} \iff \mathbf{K}_f$ is positive semidefinite and $\hat{f}_k = \hat{f}_{k'}$ if $\lambda_k = \lambda_{k'} \iff \hat{f}_k \geq 0$ and $\hat{f}_k = \hat{f}_{k'}$ if $\lambda_k = \lambda_{k'}$. Let's remember that $\mathcal{A}_{\mathrm{L}} = \tilde{A} = \left\{ f \in \mathcal{L}(G) \,\middle|\, \hat{f}_k = \hat{f}_{k'} \text{ if } \lambda_k = \lambda_{k'} \right\}$.

(2) $f \in \mathcal{B}_M = \{u_1, \ldots, u_M\}$ with $M \leq n \iff$ all the eigenvalues with index $k > M$ are zero $\iff \hat{f}_k = 0$ for $k > M$.

(3) f is conditionally positive definite with respect to the subspace $\mathcal{Y} = \mathrm{span}\,\{u_{k_1}, \ldots, u_{k_K}\}$ $\iff \mathbf{K}_f$ is positive definite on the subspace $\mathcal{Y} \iff \hat{f}_{k_1} > 0, \ldots, \hat{f}_{k_K} > 0$.

This concludes the proof. □

Although Bochner's characterization of p.s.d. functions in $\mathbb{R}^d$ is a deep result, Theorem (4.2.1) can be inferred almost directly from definition (4.1.1) for p.d. functions on graphs. This is due to the vector space of signals $\mathcal{L}(G)$ has a finite dimension, simplifying various aspects of function spaces, and the convolution in (2.7) is closely related to the spectrum of the graph G.

Furthermore, Theorem (4.2.1) establishes a one-to-one correlation between p.s.d. functions and p.s.d. kernels on graphs with a Mercer extension in terms of the Fourier basis $\{u_1, \ldots, u_n\}$. Additionally, the theorem has a second important consequence regarding the graph C^*-algebra $\mathcal{A}$, stating that the set of p.d. functions corresponds precisely to the set of positive elements in the C^*-algebra $\mathcal{A}$.

4.3 The convex cone of positive definite functions on graphs

First of all, the convex cone of positive definite functions on graphs can be defined as follows

Definition 4.3.1. ***(Convex cone)***
A set $\mathcal{P}$ is a convex cone if and only if it satisfies:

(a) Convexity: $\forall f, g \in \mathcal{P}, \forall \lambda \in [0,1] : \lambda f + (1-\lambda) g \in \mathcal{P}$

(b) Conicality: $\forall f \in \mathcal{P}, \forall \alpha > 0 : \alpha f \in \mathcal{P}$

In particular, our focus lies in characterizing both the set $\mathcal{P}$ and its subset $\mathcal{P}_+$ containing positive definite functions. For this, we first consider the norm

$$\|x\|_{\mathcal{A}'} = \sum_{k=1}^{n} |\hat{x}_k|$$

and the bounded set

$$\mathcal{P}_1 = \{f \in \mathcal{P} \mid \|f\|_{\mathcal{A}'} \leq 1\}.$$

The norm $\|\cdot\|_{\mathcal{A}'}$ indicates the norm dual to the C^*-algebra norm $\|\cdot\|_{\mathcal{A}}$. Consequently, the set $\mathcal{P}_1$ corresponds to the intersection of the unit ball in the dual algebra $\mathcal{A}'$ with the set $\mathcal{P}$ of positive semidefinite functions. In the subsequent discussion, we denote the zero signal in $\mathcal{L}(G)$ by $\bar{0} = (0, \ldots, 0)^{\mathrm{T}}$. As an initial outcome, we achieve the following characterization of the convex set $\mathcal{P}_1$.

Theorem 4.3.1. ***(Convex characterization)***
The signals $\{\bar{0}, u_1, \ldots, u_n\}$ are the extreme points of the convex and compact set $\mathcal{P}_1$. In particular,

we have

$$\mathcal{P}_1 = conv\left\{\bar{0}, u_1, \ldots, u_n\right\}. \tag{4.1}$$

Proof:

Let's consider the n-dimensional Euclidean space, denoted by $\mathbb{R}^n$, equipped with its standard basis vectors $\{e_1, e_2, \ldots, e_n\}$. In this space, the standard simplex, denoted by Δ, has vertices that include the origin $\bar{0}$ and the standard basis vectors $e_1, e_2, \ldots, e_n$. We can see that the standard simplex Δ has two key properties:

(1) **Convexity:** $\forall\, u, v \in \Delta, \forall\, \lambda \in [0,1]$: $\lambda u + (1-\lambda)v \in \Delta$

(2) **Compactness:** Δ is closed and bounded

Since $\Delta = \text{conv}\left(\{\bar{0}, u_1, \ldots, u_n\}\right) = \{\lambda \in \mathbb{R}^n \mid \sum_{k=1}^n \lambda_k = 1,\ \lambda_k \geq 0\}$ we have the following:

(1) Convexity:

Let $u = e^{\mathrm{T}}\lambda$ and $v = e^{\mathrm{T}}\mu$, and $\gamma \in [0,1]$. We have $\sum_{k=1}^n \lambda_k = \sum_{k=1}^n \mu_k = 1, \lambda_k, \mu_k \geq 0$ for all $k \in \{1, \ldots, n\}$. Let $w = \gamma u + (1-\gamma)v = \gamma e^{\mathrm{T}}\lambda + (1-\gamma)e^{\mathrm{T}}\mu = \gamma + (1-\gamma) = 1$, and $\gamma\lambda_k + (1-\gamma)\mu_k \geq 0$ given that $\gamma,\ 1-\gamma,\ \lambda_k,\ \mu_k$ are non-negative. Hence, $w \in \Delta \Rightarrow \Delta$ is convex.

(2) Compactness:

Step 1: the simplex $\Delta = \{\lambda \in \mathbb{R}^n \mid \sum_{k=1}^n \lambda_k = 1, \lambda_k \geq 0\}$ is compact. Indeed, it is closed, being the intersection of closed sets, namely $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$, and the hyperplane $H := \{\lambda \in \mathbb{R}^n \mid \sum_{k=1}^n \lambda_k = 1\}$. In addition, Δ is contained within the hypercube $[0,1]^n$, thus it is bounded. To see the last statement let's recall the definition of the hypercube $[0,1]^n$, denoted as Hyper($[0,1]^n$)= $\{(x_1, \ldots, x_n) \mid |x_i| \leq 1,\ \forall\, i = 1, \ldots, n\}$. Then $\forall\, x \in \Delta \Rightarrow \sum_{k=1}^n x_k = 1 \Rightarrow x_k \leq 1\, \forall\, k \in \{1, \ldots, n\} \Rightarrow x \in \text{Hyper}([0,1]^n) \Rightarrow$ $\Rightarrow \Delta \subset \text{Hyper}([0,1]^n)$.

Step 2: the map $g : \Delta \to \mathbb{R}^n, \lambda \to \sum_{k=1}^n \lambda_k = 1$ is continuous. Thus, conv($\{\bar{0}, u_1, \ldots, u_n\}$) is equal to $g(\Delta)$, which is the continuous image of a compact set and is therefore itself compact.

Hence, the simplex Δ is convex, compact and has precisely the $n+1$ mentioned extreme points. Now, through the natural correspondence between $\mathcal{L}(\hat{G})$ and $\mathbb{R}^n$ we can interpret Δ as a convex simplex in $\mathcal{L}(\hat{G})$ and apply the inverse Fourier transform **U**. Then, Bochner's characterization in Theorem (4.2.1) implies that:

$$\mathbf{U}\Delta = \mathcal{P}_1.$$

Since **U** is both linear and invertible (i.e., continuous), the image $\mathcal{P}_1$ of the convex set Δ is convex and compact, as illustrated in **Step 2**. Furthermore, the extreme points e_k of Δ are mapped onto the following extreme points:

$$\mathbf{U}e_k = \begin{pmatrix} \sum_{j=1}^{n} \underbrace{u_{1j} e_{k_j}}_{\delta_{jk}} \\ \vdots \\ \sum_{j=1}^{n} \underbrace{u_{nj} e_{k_j}}_{\delta_{jk}} \end{pmatrix} = \begin{pmatrix} u_{1k} \\ \vdots \\ u_{nk} \end{pmatrix} = u_k.$$

Here, u_{ij} indicates the element located at row i and column j of the matrix **U**. Similarly, e_{k_j} represents the j-th element of the vector e_k. This shows the statement of the theorem. □

We conclude this section with a list of elementary properties of the set $\mathcal{P}$ and $\mathcal{P}_1$.

Corollary 4.3.1.

(a) If $f, g \in \mathcal{P}$, then $\gamma_1 f + \gamma_2 g \in \mathcal{P}$ for all $\gamma_1, \gamma_2 \geq 0$ ($\mathcal{P}$ is a convex cone in $\mathcal{L}(G)$).

(b) The extreme radii of the cone $\mathcal{P}$ are given by $\{\gamma u_k \mid \gamma \geq 0\}$, $k \in \{1, \ldots, n\}$.

(c) The convex cone $\mathcal{P}_+$ is the open interior of $\mathcal{P}$.

(d) If $f, g \in \mathcal{P}$, then $f * g \in \mathcal{P}$ ($\mathcal{P}$ is closed under convolution).

(e) $f \in \mathcal{P}$ if and only if there exists a $g \in \mathcal{L}(G)$ with $f = g * g$.

Proof:

(a) Let $\mathbf{K}_f$ and $\mathbf{K}_g$ denote the positive semidefinite matrices associated with the functions f and g respectively. In addition, let $\gamma_1, \gamma_2 \geq 0$. Then, $\gamma_1 f + \gamma_2 g \in \mathcal{P} \iff \gamma_1 \mathbf{K}_f + \gamma_2 \mathbf{K}_g$ is positive semidefinite. Indeed $x^{\mathrm{T}} \left(\gamma_1 \mathbf{K}_f + \gamma_2 \mathbf{K}_g\right) x = \gamma_1 x^{\mathrm{T}} \mathbf{K}_f x + \gamma_2 x^{\mathrm{T}} \mathbf{K}_g x \geq 0$. Here, we have leveraged the algebraic properties of matrix multiplication along with the positive semi-definiteness of matrices $\mathbf{K}_f$ and $\mathbf{K}_g$. As a result, we can infer that $\gamma_1 f + \gamma_2 g \in \mathcal{P}$.

(b) This statement is a consequence of the Theorem (4.3.1) with the additional observation that $\mathcal{P} = \cup_{\gamma \geq 0} \gamma \mathcal{P}_1$. In order to show the observation we need to prove two points: **(i)** if $f \in \mathcal{P} \Rightarrow f \in \cup_{\gamma \geq 0} \gamma \mathcal{P}_1$; **(ii)** if $f \in \cup_{\gamma \geq 0} \gamma \mathcal{P}_1 \Rightarrow f \in \mathcal{P}$.

- **(i)** $f \in \mathcal{P} \Rightarrow f \in \mathcal{L}(G)$ such that f is p.s.d. $\Rightarrow$ the kernel matrix $\mathbf{K}_f$ is p.s.d. $\Rightarrow$ all the eigenvalues of $\mathbf{K}_f$ are non-negative, i.e. $\hat{f}_k \geq 0 \, \forall \, k \in \{1, \ldots, n\}$ due to Theorem (4.2.1) $\Rightarrow \exists \, \gamma \geq 0$ such that $\gamma f \in \mathcal{P}$ and $\gamma \sum_{k=1}^{n} |\hat{f}_k| \leq 1$; for example we can take $\gamma = \frac{1}{1 + \sum_{k=1}^{n} |\hat{f}_k|} \Rightarrow \gamma f \in \mathcal{P}$ and $\|\gamma f\|_{\mathcal{A}'} \leq 1 \Rightarrow \exists \, \gamma \geq 0$ such that $f \in \gamma \mathcal{P}_1 \Rightarrow f \in \cup_{\gamma \geq 0} \gamma \mathcal{P}_1$.

- **(ii)** $f \in \cup_{\gamma \geq 0} \gamma \mathcal{P}_1 \Rightarrow \exists \, \gamma \geq 0$ such that $f \in \gamma \mathcal{P}_1 \Rightarrow \exists \, \gamma \geq 0$ such that $\gamma f \in \mathcal{P}$ and $\|\gamma f\|_{\mathcal{A}'} \leq 1 \Rightarrow$ since $\gamma f \in \mathcal{P}$ we have that $\widehat{\gamma f_k} = \mathbf{U}^{\mathrm{T}} \gamma f_k = \gamma \hat{f}_k \geq 0$, due to Theorem (4.2.1) $\Rightarrow$ now, given that $\gamma \geq 0$ it follows that $\hat{f}_k$ must be non-negative for all $k \in \{1, \ldots, n\}$, i.e. $\hat{f}_k \geq 0 \, \forall \, k \in \{1, \ldots, n\} \Rightarrow f \in \mathcal{P}$, consistently as for the Theorem (4.2.1)

Currently, given that $\mathcal{P} = \cup_{\gamma \geq 0} \gamma \mathcal{P}_1$, and $\mathcal{P}_1 = \mathbf{U}\Delta$, $\mathbf{U}e_k = u_k$ from the Theorem (4.3.1), we can conclude the statement in the following way:
$\text{conv}(\mathcal{P}) = \text{conv}(\cup_{\gamma \geq 0} \gamma \mathcal{P}_1) = \text{conv}(\cup_{\gamma \geq 0} \gamma \mathbf{U}\Delta) = \text{conv}(\cup_{\gamma \geq 0} \gamma \mathbf{U}\Delta) =$
$= \text{conv}(\cup_{\gamma \geq 0} \gamma \{\bar{0}, u_1, \ldots, u_n\}) = \cup_{\gamma \geq 0} \text{conv}(\{\bar{0}, \gamma u_1, \ldots, \gamma u_n\}) \Rightarrow$ the extreme radii of the cone $\mathcal{P}$ are $\{\gamma u_k \mid \gamma \geq 0\}$ for all $k \in \{1, \ldots, n\}$.

(c) By Bochner's characterization in Theorem (4.2.1) a positive semidefinite function f is on the boundary of the cone $\mathcal{P}$ if and only if $\hat{f}_k = 0$ for at least one $k \in \{1, \ldots, n\}$. On the other hand, this is equivalent for f to be in $\mathcal{P} \backslash \mathcal{P}_+$. To understand that the convex cone $\mathcal{P}_+$ is the open interior of $\mathcal{P}$, let's consider a set V with a function $f : V \to \mathbb{R}$. The boundary of V, denoted by ∂V, is defined as the set of points in V that do not have a neighborhood completely contained in V, that is, the points that are "on the boundary" between the interior and exterior of V. If a function f is not in the boundary of V, it means that for every point v in the domain of f, there exists a neighborhood of v that is completely contained in V. This implies that $f(v)$ is in the interior of V, because it's possible to "move around" a bit in all directions without leaving the set V. So, if a function is not in the boundary of its set, then it certainly lies in its interior. The fact that $\mathcal{P}_+$ is a convex cone is also guaranteed by definition (4.1.1)

(d) We can observe that $f * g \in \mathcal{P}$ if and only if $(\widehat{f * g})_k \geq 0$ for all $k \in \{1, \ldots, n\}$. Thus, we can proceed as follows: $(\widehat{f * g})_k = (\mathbf{U}^{\mathrm{T}} f * g)_k = (\mathbf{U}^{\mathrm{T}} \mathbf{U} \mathbf{M}_{\hat{f}} \hat{g})_k = (\hat{f}_1 \hat{g}_1, \ldots, \hat{f}_n \hat{g}_n)_k = \hat{f}_k \hat{g}_k \geq 0$ given that $f, g \in \mathcal{P}$, that is, $\hat{f}_k, \hat{g}_k \geq 0 \, \forall \, k \in \{1, \ldots, n\} \Rightarrow (\widehat{f * g})_k \geq 0 \, \forall \, k \in \{1, \ldots, n\} \Rightarrow f * g \in \mathcal{P}$.

(e) $f \in \mathcal{P} \iff \exists\, g \in \mathcal{L}(G)$ such that $f = g * g$.
$(\Longrightarrow)$ We know that $f \in \mathcal{P}$, then f can be expressed as $f = \sum_{k=1}^{n} \hat{f}_k u_k$, with $\hat{f}_k \geq 0\, \forall\, k \in \{1, \dots, n\}$. Now, let $g \in \mathcal{L}(G)$. Therefore, the convolution with itself is given by $g * g = \mathbf{U}^{\mathrm{T}} \mathbf{M}_{\hat{g}} \hat{g} = \sum_{k=1}^{n} \hat{g}_k^2 u_k$. Since $g = \mathbf{U}\sqrt{\mathbf{U}^{\mathrm{T}} f}$, $\hat{g}_k^2 = (\widehat{\mathbf{U}\sqrt{\mathbf{U}^{\mathrm{T}} f}})_k^2 = (\mathbf{U}^{\mathrm{T}}\mathbf{U}\sqrt{\mathbf{U}^{\mathrm{T}} f})_k^2 = (\mathbf{U}^{\mathrm{T}} f)_k = \hat{f}_k \Rightarrow g * g = \sum_{k=1}^{n} \hat{f}_k u_k = f$.
$(\Longleftarrow)$ We know that $\exists\, g \in \mathcal{L}(G)$ such that $f = g * g$. Thus, $\sum_{k=1}^{n} \hat{f}_k u_k = f = g * g = \sum_{k=1}^{n} \hat{g}_k^2 u_k \Rightarrow$ the chain of equalities holds true if and only if $\hat{f}_k = \hat{g}_k^2\, \forall\, k \in \{1, \dots, n\} \Rightarrow$ since $\hat{g}_k^2 \geq 0\, \forall\, k \in \{1, \dots, n\}$, inevitably, also $\hat{f}_k \geq 0\, \forall\, k \in \{1, \dots, n\} \Rightarrow f \in \mathcal{P}$. □

Theorem (4.3.1) also suggests that we can express any positive semidefinite function $f \in \mathcal{P}_1$ as:

$$f = \lambda_1 u_1 + \ldots + \lambda_n u_n,\ \ \lambda_k \geq 0,\ \ \sum_{k=1}^{n} \lambda_k \leq 1.$$

As $\mathcal{P}_+$ represents the open interior of $\mathcal{P}$, any $f \in \mathcal{P}_+ \cap \mathcal{P}_1$ must intrinsically take the following form:

$$f = \lambda_1 u_1 + \ldots + \lambda_n u_n,\ \ \lambda_k > 0,\ \ \sum_{k=1}^{n} \lambda_k \leq 1.$$

Here, $\lambda_k,\ u_k$ denote the eigenvalues and eigenvectors, respectively, of the associated kernel matrix $\mathbf{K}_f$. Now, in this scenario we show that $\sum_{k=1}^{n} \lambda_k \leq 1$.

Proof:
Given that $f \in \mathcal{P}_1$ or $f \in \mathcal{P}_+ \cap \mathcal{P}_1$ we can infer that $|\lambda_k| = \lambda_k\, \forall\, k \in \{1, \dots, n\}$ by Theorem (4.2.1). Furthermore, the following chain of equivalences holds:

$$\|f\|_{\mathcal{A}'} \leq 1 \iff \sum_{k=1}^{n} \left|\hat{f}_k\right| \leq 1 \iff \sum_{k=1}^{n} \left|\mathbf{U}^{\mathrm{T}} f_k\right| \leq 1 \iff \sum_{k=1}^{n} \left|\mathbf{U}^{\mathrm{T}} \lambda_k u_k\right| \leq 1 \iff$$
$$\iff \sum_{k=1}^{n} \lambda_k \left|\mathbf{U}^{\mathrm{T}} u_k\right| \leq 1 \iff \sum_{k=1}^{n} \lambda_k \underbrace{|e_k|}_{1} \leq 1 \iff \sum_{k=1}^{n} \lambda_k \leq 1.$$

Thus, the inequality is shown. □

4.4 Moment conditions for positive definite functions

Let's start this section by recalling the definition of a Hankel matrix.

Definition 4.4.1. ***(Hankel Matrix)***
A Hankel matrix is a square matrix $H = [h_{ij}]$ of size $n \times n$ with elements h_{ij} satisfying the condition $h_{ij} = h_{i+1,j+1}$ for all i, j such that $i+1$ and $j+1$ are within the bounds of the matrix dimensions, i.e., $1 \leq i, j \leq n-1$.

We consider the moments $f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^j x,\ j \in \mathbb{N}_0$, of a signal $x \in \mathcal{L}(G)$. With these moments we can generate the Hankel matrices

$$\mathbf{H}_r(x) = \begin{pmatrix} f_{\mathbb{1}}^{\mathrm{T}}x & f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}x & \cdots & f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^{r-1}x \\ f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}x & f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^2x & \cdots & f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^{r}x \\ \vdots & \vdots & \ddots & \vdots \\ f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^{r-1}x & f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^{r}x & \cdots & f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^{2r-2}x \end{pmatrix}, \quad r \in \mathbb{N}. \tag{4.2}$$

The moment matrices $\mathbf{H}_r(x)$ allow to characterize p.d. functions in the subalgebra $\mathcal{A}_{\mathbf{L}}$.

Theorem 4.4.1.
Assume that the normalized graph Laplacian $\mathbf{L}$ has exactly r distinct eigenvalues. A signal $f \in \mathcal{A}_L$ is positive (semi-) definite if and only if the matrix $\mathbf{H}_r(f)$ is positive (semi-) definite.

Proof:
First of all, we show that if $f \in \mathcal{P}$ then the matrix $\mathbf{H}_r(f)$ is positive semi-definite. For an arbitrary vector $y = (y_1, \ldots, y_n)^{\mathrm{T}} \in \mathbb{R}^n$, the following identities hold:

$$\begin{aligned} y^{\mathrm{T}}\mathbf{H}_r(f)y &= \sum_{i=1}^{r}\sum_{j=1}^{r} f_{\mathbb{1}}^{\mathrm{T}}\mathbf{L}^{i+j-2}f y_i y_j \overset{(2.2)}{=} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} f_{\mathbb{1}}^{\mathrm{T}}\mathbf{U}\mathbf{M}_{\lambda}^{i+j-2}\underbrace{\mathbf{U}^{\mathrm{T}}f}_{\hat{f}}\, y_i y_j = \\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} f_{\mathbb{1}}^{\mathrm{T}}\mathbf{U}\mathbf{M}_{\lambda}^{i+j-2}\hat{f} y_i y_j = \ldots \end{aligned}$$

Note that $f_{\mathbb{1}}^{\mathrm{T}}\mathbf{U} = [\underbrace{(u_1^{\mathrm{T}} + \ldots + u_n^{\mathrm{T}})u_1}_{\delta_{ij}}, \ldots, \underbrace{(u_1^{\mathrm{T}} + \ldots + u_n^{\mathrm{T}})u_n}_{\delta_{ij}}] = (1, \ldots, 1)$, then

$$\ldots = \sum_{i=1}^{r}\sum_{j=1}^{r}(1, \ldots, 1)\mathbf{M}_{\lambda}^{i+j-2}\hat{f}y_i y_j = \sum_{i=1}^{r}\sum_{j=1}^{r}(\lambda_1^{i+j-2}, \ldots, \lambda_n^{i+j-2})\hat{f}y_i y_j =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{n}\hat{f}_k\lambda_k^{i+j-2}y_i y_j = \sum_{k=1}^{n}\hat{f}_k\sum_{i=1}^{r}\sum_{j=1}^{r}\lambda_k^{i+j-2}y_i y_j = \ldots \tag{4.3}$$

Let's take a closer look at the term $\sum_{i=1}^{r}\sum_{j=1}^{r}\lambda_k^{i+j-2}y_i y_j$, thus

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\lambda_k^{i+j-2}y_i y_j = y^{\mathrm{T}}\begin{pmatrix} \lambda_k^0 & \lambda_k^1 & \ldots & \lambda_k^{r-1} \\ \lambda_k^1 & \lambda_k^2 & \ldots & \lambda_k^r \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_k^{r-1} & \lambda_k^r & \ldots & \lambda_k^{2r-2} \end{pmatrix} y =$$

$$= y^{\mathrm{T}}\begin{pmatrix} \lambda_k^0 \\ \vdots \\ \lambda_k^{r-1} \end{pmatrix}(\lambda_k^0, \ldots, \lambda_k^{r-1})y = y^{\mathrm{T}}(\lambda_k^0, \ldots, \lambda_k^{r-1})^{\mathrm{T}}(\lambda_k^0, \ldots, \lambda_k^{r-1})y =$$

$$= \left[(\lambda_k^0, \ldots, \lambda_k^{r-1})y\right]^{\mathrm{T}}\left[(\lambda_k^0, \ldots, \lambda_k^{r-1})y\right] = \left[(\lambda_k^0, \ldots, \lambda_k^{r-1})y\right]^2 \tag{4.4}$$

Now, plugging (4.4) into (4.3), we obtain

$$y^{\mathrm{T}}\mathbf{H}_r(f)y = \sum_{k=1}^{n}\hat{f}_k\underbrace{\left[(\lambda_k^0, \ldots, \lambda_k^{r-1})y\right]^2}_{\geq 0}. \tag{4.5}$$

Therefore, if f is positive semidefinite, Theorem (4.2.1) implies that $\mathbf{H}_r(f)$ is also positive semidefinite. To establish the opposite conclusion, we need the r distinct eigenvalues of the graph Laplacian $\mathbf{L}$, denoted by $\lambda_{k_1}, \ldots, \lambda_{k_r}$. Since these eigenvalues are distinct, the Vandermonde matrix

$$\mathbf{V}_r = \begin{pmatrix} \lambda_{k_1}^0 & \lambda_{k_1}^1 & \ldots & \lambda_{k_1}^{r-1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k_r}^0 & \lambda_{k_1}^1 & \ldots & \lambda_{k_r}^{r-1} \end{pmatrix}$$

is invertible (see, (2.8)). Hence, for each $j \in \{1, \ldots, r\}$ there exists a unique vector $y^{(j)} \in \mathbb{R}^r,\ y^{(j)} \neq \bar{0}$ such that $\mathbf{V}_r y^{(j)} = e_j \in \mathbb{R}^r$. Substituting these solutions $y^{(j)}$ into (4.5), we obtain

$$
\begin{aligned}
y^{\mathrm{T}}\mathbf{H}_r(f)y = \sum_{k=1}^{n} \hat{f}_k \quad & \underbrace{\left[(\lambda_k^0, \ldots, \lambda_k^{r-1})y^{(j)}\right]^2}_{\mathbf{V}_r y^{(j)} = e_j,\, k\colon \lambda_k = \lambda_{k_j},\, j \in \{1,\ldots,r\}} \quad = \\
= \sum_{k:\lambda_k=\lambda_{k_j}} \hat{f}_k(e_j)^2 = \sum_{k:\lambda_k=\lambda_{k_j}} \hat{f}_k \underbrace{e_j^{\mathrm{T}} e_j}_{1} = \sum_{k:\lambda_k=\lambda_{k_j}} \hat{f}_k \,.
\end{aligned}
$$

Thus, assuming that $\mathbf{H}_r(f)$ is positive semidefinite and utilizing the characterization of the subalgebra $\mathcal{A}_{\mathbf{L}}$ in Proposition (2.4.1), we infer that $\hat{f}_k \geq 0$ for all $k \in \{1,\ldots,n\}$. Consequently, this implies that f is a positive semidefinite function. For the stricter assumption that $f \in \mathcal{P}_+$, the reasoning in the proof is almost the same. The only discrepancy is in showing the forward direction ($f \in \mathcal{P}_+$) $\Rightarrow$ ($\mathbf{H}_r(f)$ is positive definite). In this scenario, the requirement of having exactly r distinct eigenvalues for $\mathbf{L}$ is already necessary. □

5

Interpolation With Graph Basis Functions

Using the relation $K_f(v_i, v_j) = \mathbf{C}_{e_j} f(v_i)$, a positive definite function f generates a positive definite kernel K_f. Then, an interpolation scheme for the generalized translates $\mathbf{C}_{e_j} f$ of the graph basis function f is obtained by employing the corresponding kernel-based approach outlined in Chapter (3). We summarize this interpolation scheme in the **Algorithm**.

Algorithm: Interpolation with Graph Basis Functions (GBF's)

Input: Signal values $x(w_1), \ldots, x(w_N)$ at the sampling nodes $W \subset V$. A positive definite graph basis function $f \in \mathcal{P}^+$.
Calculate the N generalized translates $\mathbf{C}_{e_{j_1}} f = e_{j_1} * f, \ldots, \mathbf{C}_{e_{j_N}} f = e_{j_N} * f$ with the correspondence $v_{j_k} = w_k$ for the nodes in W.
Solve the linear system of equations

$$\begin{pmatrix} \mathbf{C}_{e_{j_1}} f(w_1) & \mathbf{C}_{e_{j_2}} f(w_1) & \cdots & \mathbf{C}_{e_{j_N}} f(w_1) \\ \mathbf{C}_{e_{j_1}} f(w_2) & \mathbf{C}_{e_{j_2}} f(w_2) & \cdots & \mathbf{C}_{e_{j_N}} f(w_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{e_{j_1}} f(w_N) & \mathbf{C}_{e_{j_2}} f(w_N) & \cdots & \mathbf{C}_{e_{j_N}} f(w_N) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} = \begin{pmatrix} x(w_1) \\ x(w_2) \\ \vdots \\ x(w_N) \end{pmatrix}.$$

Calculate the GBF-interpolant

$$\mathbf{I}_W x(v) = \sum_{k=1}^{N} c_k \mathbf{C}_{e_{j_k}} f(v).$$

In the following, we refer to $\mathbf{I}_W x$ as the GBF interpolant of the signal x on the nodes W. The specific structure of this interpolation scheme enables us to discuss error estimates and stability issues in Chapter (8) much like RBF interpolation in $\mathbb{R}^d$ or SBF interpolation on the unit sphere. The GBF interpolation space is defined by the generalized translates $\mathbf{C}e_{j_k} f$ of the GBF f, i.e.,

$$\mathcal{N}_{K_f,W} = \left\{ x \in \mathcal{L}(G) \mid x = \sum_{k=1}^{N} c_k \mathbf{C}_{e_{j_k}} f \right\}.$$

Bochner's characterization in Theorem (4.2.1) provides us the following characterization of the inner product within the native space $\mathcal{N}_{K_f} = \mathcal{N}_{K_f,V}$.

Theorem 5.0.1.
if $f \in \mathcal{P}_+$ then the inner product and the norm of the native space $\mathcal{N}_{K_f}$ are given as

$$\langle x, y \rangle_{K_f} = \sum_{k=1}^{n} \frac{\hat{x}_k \hat{y}_k}{\hat{f}_k} = \hat{y}^T \boldsymbol{M}_{1/\hat{f}} \hat{x} \;\; and \;\; \|x\|_{K_f} = \sqrt{\sum_{k=1}^{n} \frac{\hat{x}_k^2}{\hat{f}_k}}.$$

Proof:
According to the characterization in Theorem (4.2.1), the eigendecomposition of the positive definite matrix $\mathbf{K}_f$ can be expressed as $\mathbf{K}_f = \mathbf{U}\mathbf{M}_{\hat{f}}\mathbf{U}^{\mathrm{T}}$. Consequently, the inverse $\mathbf{K}_f^{-1}$ of $\mathbf{K}_f$ takes the form $\mathbf{K}_f^{-1} = \mathbf{U}\mathbf{M}_{1/\hat{f}}\mathbf{U}^{\mathrm{T}}$. Moreover, the inner product of the native space $\mathcal{N}_{K_f}$ can be written as

$$\langle x, y \rangle_{K_f} = y^{\mathrm{T}} \mathbf{K}_f^{-1} x = \underbrace{y^{\mathrm{T}} \mathbf{U}}_{(\mathbf{U}^{\mathrm{T}} y)^{\mathrm{T}} = \hat{y}^{\mathrm{T}}} \mathbf{M}_{1/\hat{f}} \underbrace{\mathbf{U}^{\mathrm{T}} x}_{\hat{x}} = \hat{y}^{\mathrm{T}} \mathbf{M}_{1/\hat{f}} \hat{x} = \sum_{k=1}^{n} \frac{\hat{x}_k \hat{y}_k}{\hat{f}_k},$$

$$\|x\|_{K_f}^2 = \langle x, x \rangle_{K_f} = \hat{x}^{\mathrm{T}} \mathbf{M}_{1/\hat{f}} \hat{x} = \sum_{k=1}^{n} \frac{\hat{x}_k^2}{\hat{f}_k} \Longrightarrow \|x\|_{K_f} = \sqrt{\sum_{k=1}^{n} \frac{\hat{x}_k^2}{\hat{f}_k}}.$$

Thus, the proof of the theorem is complete. □

6

Examples Of Positive Definite Functions on Graphs

Below, we outline several significant examples of positive definite graph basis functions (GBF's). Some of these examples are to well-known graph kernels.

(1) Trivial interpolation with the unity $f_{\mathbb{1}}$: The unity $f_{\mathbb{1}} = \sum_{k=1}^{n} u_k$ of the graph convolution is a positive definite function. We have $\mathbf{C}_{e_i} f_{\mathbb{1}} = e_i$. Therefore, $\mathbf{K}_{f_{\mathbb{1}}} = \mathbf{I}_n$ and $\mathbf{K}_{f_{\mathbb{1}},W} = \mathbf{I}_N$ are identity matrices and the interpolation space of the GBF $f_{\mathbb{1}}$ is given by $\mathcal{N}_{\mathbf{K}_{f_{\mathbb{1}}},W} = \text{span}\,\{e_{j_1}, \ldots, e_{j_N}\}$ (here, v_{j_k} corresponds to the node w_k). Interpolation in terms of the basis functions $\{e_{j_1}, \ldots, e_{j_N}\}$ is nothing else than extending the given data $x(w_1), \ldots, x(w_N)$ by $x(v) = 0$ for all $v \in V \backslash W$.
Let's show that $\mathbf{C}_{e_i} f_{\mathbb{1}} = e_i$:

$$\begin{aligned}\mathbf{C}_{e_i} f_{\mathbb{1}} &= \mathbf{U}\mathbf{M}_{\hat{e}_i}\mathbf{U}^{\mathrm{T}} f_{\mathbb{1}} = \mathbf{U}\text{diag}\left(\mathbf{U}^{\mathrm{T}} e_i\right)\mathbf{U}^{\mathrm{T}} f_{\mathbb{1}} = \\ &= \mathbf{U}\mathbf{U}^{\mathrm{T}}\text{diag}(e_i)\mathbf{U}^{\mathrm{T}} f_{\mathbb{1}} = \text{diag}(e_i)\mathbf{U}^{\mathrm{T}} f_{\mathbb{1}} = \\ &= \text{diag}(e_i)(1, \ldots, 1)^{\mathrm{T}} = e_i.\end{aligned}$$

(2) The graph Laplacian L: A prominent example is the graph Laplacian $\mathbf{L}$ for a connected graph G. Since the eigenvalues λ_k of $\mathbf{L}$ are all positive except for $\lambda_1 = 0$, the graph Laplacian $\mathbf{L}$ corresponds to a conditionally positive definite kernel that allows the Mercer decomposition

$$\mathbf{L} = \sum_{k=2}^{n} \lambda_k u_k u_k^{\mathrm{T}}.$$

The Laplacian $\mathbf{L}$ is positive definite on the subspace span $\{u_2, \ldots, u_n\}$ and maps the constant signals span $\{u_1\}$ to the zero signal $\bar{0}$. Therefore, for every $\delta > 0$ the augmented Laplacian

$$\mathbf{L}^{(\delta)} = \mathbf{L} + \delta u_1 u_1^{\mathrm{T}} = \sum_{k=1}^{n} \lambda_k u_k u_k^{\mathrm{T}}, \text{ with } \lambda_1 = \delta,$$

is positive definite. The positive definite generator $f_{\mathbf{L}^{(\delta)}}$ of the augmented Laplacian is determined by

$$\hat{f}_{\mathbf{L}^{(\delta)}} = (\lambda_1, \ldots, \lambda_n) = (\delta, \lambda_2, \ldots, \lambda_n)$$

(3) **Polynomials of the graph Laplacian L:** The spectral calculus allows us to define further positive definite kernels based on the eigenvalue decomposition of $\mathbf{L}$. If p_r is a positive polynomial of degree r on the interval $[0, 2]$, we get $p_r(\lambda_k) > 0$ for all eigenvalues λ_k of $\mathbf{L}$. Therefore,

$$p_r(\mathbf{L}) = \sum_{k=1}^{n} p_r(\lambda_k) u_k u_k^{\mathrm{T}}$$

gives rise to a positive definite kernel on G. The Fourier transform of the generating GBF $f_{p_r(\mathbf{L})}$ is given as

$$\hat{f}_{p_r(\mathbf{L})} = (p_r(\lambda_1), \ldots, p_r(\lambda_n)).$$

These positive definite GBF's are relevant for practical applications, especially if the size n of G gets large. In this case, the kernel matrix $p_r(\mathbf{L})$ and its columns can be calculated quickly with simple matrix-vector multiplications based on the graph Laplacian $\mathbf{L}$.
Observation:
Define the following functions f in terms of the Fourier transform:

$$\hat{f}_k = \sum_{l=0}^{r} \alpha_l \lambda_k^l = p_r(\lambda_k), \;\; r \leq n.$$

Then we have

$$(f * x) = \mathbf{U}\mathrm{diag}(\hat{f})\mathbf{U}^{\mathrm{T}}x = \mathbf{U}\mathrm{diag}(p_r(\lambda_1), \dots, p_r(\lambda_n))\mathbf{U}^{\mathrm{T}}x =$$

$$p_r(\underbrace{\mathbf{U}\mathrm{diag}(\lambda_1, \dots, \lambda_n)\mathbf{U}^{\mathrm{T}}}_{\mathbf{L}})x = p_r(\mathbf{L})x = \sum_{k=0}^{r} \alpha_k \mathbf{L}^k x.$$

(4) Variational or polyharmonic splines: Variational splines on graphs were first introduced in [13] as solutions denoted by $\mathbf{I}_W x$ to the interpolation problem (3.2). These solutions minimize the functional $\|(\varepsilon \mathbf{I}_n + \mathbf{L})^{s/2}\mathbf{I}_n x\|$, where $\varepsilon > 0$ and $s > 0$. According to Proposition (3.3.4), this energy functional corresponds to the native space norm of the kernel

$$(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s} = \sum_{k=1}^{n} \frac{1}{(\varepsilon + \lambda_k)^s} u_k u_k^{\mathrm{T}}.$$

Therefore, variational spline interpolation can be viewed as a GBF interpolation approach relying on the positive definite function $f_{(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s}}$ defined in the spectral domain as

$$\hat{f}_{(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s}} = \left(\frac{1}{(\varepsilon + \lambda_1)^s}, \dots, \frac{1}{(\varepsilon + \lambda_n)^s} \right).$$

In [14], the parameter choice $\varepsilon = 0$ was also considered. In this scenario, the functional $\|\mathbf{L}^{s/2}\mathbf{I}_W x\|$ represents a semi-norm related to the positive semidefinite kernel $(\mathbf{L}^{\dagger})^{-s}$. When the graph G is connected, the Fourier transform of the positive semidefinite function $f(\mathbf{L}^{\dagger})^{-s}$ is expressed

$$\hat{f}_{(\mathbf{L}^{\dagger})^{-s}} = \left(0, \lambda_2^{-s}, \dots, \lambda_n^{-s}\right).$$

The GBF $f(\mathbf{L}^{\dagger})^{-s}$ is therefore a conditionally positive definite function concerning the subspace span$\{u_2, \dots, u_n\}$. To ensure uniqueness, the interpolation problem can be treated as in Example **(2)** or as described in Section (3.4). A more conventional method for solving the interpolation problem with variational splines is outlined in [14], or alternatively in [37] for related issues on manifolds.

Observation 1:

We aim to show that $((\lambda_k + \varepsilon)^{-s}, u)$ is an eigenpair of $(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s}$. We will carry out the observation by using induction for a generic $\lambda_k + \varepsilon$:

Step 1: $s = 1$

$$(\varepsilon \mathbf{I}_n + \mathbf{L})^{-1}u = (\lambda_k + \varepsilon)^{-1}u \iff (\varepsilon \mathbf{I}_n + \mathbf{L})u = (\lambda_k + \varepsilon)u \iff \mathbf{L}\lambda_k = \lambda_k u$$

Step 2: Now, we assume that the assertion is true for s and we show the statement for $s+1$

$$(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s-1} u = (\varepsilon \mathbf{I}_n + \mathbf{L})^{-1} \underbrace{(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s} u}_{(\lambda_k+\varepsilon)^{-s}u} = \\ = (\lambda_k + \varepsilon)^{-s} (\varepsilon \mathbf{I}_n + \mathbf{L})^{-1} u = \\ = (\lambda_k + \varepsilon)^{-s} (\lambda_k + \varepsilon)^{-1} u = (\lambda_k + \varepsilon)^{-s-1} u.$$

Observation 2:
We want to show that (λ_k^s, u) is an eigenpair of $\mathbf{L}^s$

$$\mathbf{L}^s u = \mathbf{L}^{s-1} \mathbf{L} u = \mathbf{L}^{s-1} \lambda_k u = \ldots = \lambda_k^s u.$$

(5) Diffusion kernels: Diffusion kernels [10] based on the Mercer decomposition

$$e^{-t\mathbf{L}} = \sum_{k=1}^{n} e^{-t\lambda_k} u_k u_k^{\mathrm{T}},$$

are as well positive definite for all $t \in \mathbb{R}$. The graph Fourier transform of the respective positive definite function $f_{e^{-t\mathbf{L}}}$ is given as

$$\hat{f}_{e^{-t\mathbf{L}}} = (e^{-t\lambda_1}, \ldots, e^{-t\lambda_n}).$$

Observation:
Let's show that $e^{-t\mathbf{L}} = \sum_{k=1}^{n} e^{-t\lambda_k} u_k u_k^{\mathrm{T}}$:

$$e^{-t\mathbf{L}} = e^{-t\mathbf{U}\mathbf{M}_\lambda \mathbf{U}^{\mathrm{T}}} = \sum_{i=0}^{\infty} (-t)^i \frac{(\mathbf{U}\mathbf{M}_\lambda \mathbf{U}^{\mathrm{T}})^i}{i!} = \\ = \mathbf{U} \left(\sum_{i=0}^{\infty} \frac{(-t)^i \mathrm{diag}(\lambda_1^i, \ldots, \lambda_n^i)}{i!} \right) \mathbf{U}^{\mathrm{T}} = \\ = \mathbf{U} \left(\sum_{i=0}^{\infty} \mathrm{diag} \left(\frac{(-t\lambda_1)^i}{i!}, \ldots, \frac{(-t\lambda_n)^i}{i!} \right) \right) \mathbf{U}^{\mathrm{T}} = \\ = \mathbf{U}\mathrm{diag}(e^{-t\lambda_1}, \ldots, e^{-t\lambda_n}) \mathbf{U}^{\mathrm{T}} = \sum_{k=1}^{n} e^{-t\lambda_k} u_k u_k^{\mathrm{T}}. \quad (6.1)$$

Furthermore, from (6.1) we can infer that $e^{-t\lambda_k}$ represent the k-th eigenvalue.

(6) Kernels with polynomial Fourier decay: Positive definite functions $f_{\mathrm{pol},s}$ with a polynomial decay on the spectrum $\hat{G}$ are determined by the Fourier coefficients

$$\hat{f}_{\mathrm{pol},s} = \left(1, \frac{1}{2^s}, \frac{1}{3^s}, \ldots, \frac{1}{n^s}\right), \quad s > 0.$$

The corresponding kernel has the form

$$\mathbf{K}_{f_{\mathrm{pol},s}} = \sum_{k=1}^{n} \frac{1}{k^s} u_k u_k^{\mathrm{T}}.$$

These positive definite functions are relevant for us in the discussion of error estimates for GBF interpolation.

(7) **Bandlimited interpolation:** Interpolation within the space $\mathcal{B}_M$ of bandlimited functions can be illustrated employing the positive semidefinite function $f_{\mathbf{B}_M} = \sum_{k=1}^{M} u_k$, i.e., the unity element of the subalgebra $\mathcal{B}_M$. The corresponding kernel $\mathbf{B}_M = \sum_{k=1}^{M} u_k u_k^{\mathrm{T}}$ denotes the orthogonal projection onto the space $\mathcal{B}_M$. The fact that $f_{\mathbf{B}_M}$ is not strictly positive definite already suggests the lack of unisolvency for the interpolation problem (3.2) within the native space $\mathcal{N}_{K_{f_{\mathbf{B}_M}}} = \mathcal{B}_M$. This observation is also highlighted in [5], [8].

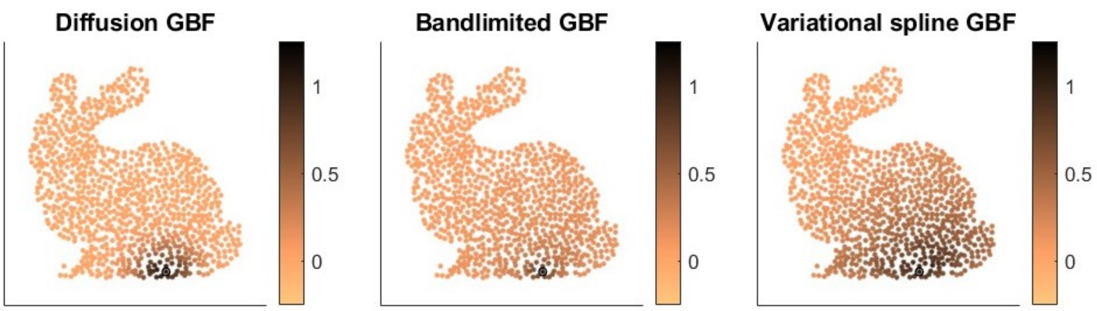


Figure 6.1: Illustration of the shifts $\mathbf{C}_{e_j} f$ for different GBF's. Left: $f = f_{e^{-10\mathbf{L}}}$ in Example **(5)**. Middle: $f = f_{\mathrm{pol},1}$ in Example **(6)**. Right: $f = f_{(0.02\mathbf{I}_n+\mathbf{L})^{-2}}$ in Example **(4)**. The ringed node corresponds to v_j.

7

Space Frequency Analysis With Positive Definite Functions

To begin this chapter, let's review some relevant facts that will help us in our next discussion.

Definition 7.0.1. ***(Spectral norm)***
Given a square matrix $A \in \mathbb{R}^{n\times n}$ the spectral norm $\|A\|_2$ is defined as

$$\|A\|_2 = \max_{x\neq\bar{0}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2.$$

Proposition 7.0.1.
Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ two square matrices, then the following inequality holds

$$\|AB\|_2 \leq \|A\|_2\|B\|_2.$$

Proof:

$$\begin{aligned}\|AB\|_2 &= \max_{x\neq\bar{0}} \frac{\|ABx\|_2}{\|x\|_2} \leq \max_{Bx\neq\bar{0}} \frac{\|ABx\|_2}{\|x\|_2}\frac{\|Bx\|_2}{\|Bx\|_2} = \\ &= \left(\max_{Bx\neq\bar{0}} \frac{\|ABx\|_2}{\|Bx\|_2}\right)\left(\max_{x\neq\bar{0}} \frac{\|Bx\|_2}{\|x\|_2}\right) = \|A\|_2\|B\|_2. \\ &\Rightarrow \|AB\|_2 \leq \|A\|_2\|B\|_2\end{aligned}$$

Thus, the inequality is shown. $\square$

Proposition 7.0.2.
Assume $u_1, \ldots, u_n$ form an orthonormal basis of $\mathbb{R}^n$. Then, we can state:

$$\sum_{k=1}^{n} u_k u_k^T = \boldsymbol{I}_n.$$

Proof:
For instance take $w = (1, 0, \ldots, 0)^{\mathrm{T}}$, then

$$w = \sum_{k=1}^{n} \langle w, u_k \rangle u_k =$$

$$= \sum_{k=1}^{n} u_k(1) u_k = \underbrace{\begin{pmatrix} \sum_{k=1}^{n} u_k(1)u_k(1) \\ \vdots \\ \sum_{k=1}^{n} u_k(1)u_k(n) \end{pmatrix}}_{\text{this is the first column of the matrix } \mathbf{I}_n} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Where $u_k(j)$ denotes the j-th entry of the vector u_k.
Now, let's consider $w^i = (0, \ldots, 0, 1, 0, \ldots, 0)$, with 1 in the i-th position, and repeat the previous computations to determine that the i-th column of $\sum_{k=1}^{n} u_k u_k^{\mathrm{T}}$ is equals to $(0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$. Therefore, this prove the equality. $\square$

Proposition 7.0.3.
By Bochner-type characterization of positive definite functions in Theorem (4.2.1) *we know that $\boldsymbol{K}_f = \boldsymbol{U}\boldsymbol{M}_{\hat{f}}\boldsymbol{U}^T$. Thus, the following inequalities holds for all $x \in \mathbb{R}^n$*

$$\left(\min_{1 \le k \le n} \hat{f}_k\right) \|x\|_2 \overset{(a)}{\le} \|\boldsymbol{K}_f x\|_2 \overset{(b)}{\le} \left(\max_{1 \le k \le n} \hat{f}_k\right) \|x\|_2$$

Proof:
(a)

$$\|\mathbf{K}_f x\|_2 = \sqrt{\langle \mathbf{K}_f x, \mathbf{K}_f x \rangle} = \sqrt{x^{\mathrm{T}} \mathbf{U} \mathbf{M}_{\hat{f}}^2 \mathbf{U}^{\mathrm{T}} x} =$$

$$= \left[x^{\mathrm{T}} \left(\sum_{k=1}^{n} \hat{f}_k^2 u_k u_k^{\mathrm{T}} \right) x \right]^{\frac{1}{2}} \geq \left[x^{\mathrm{T}} \left(\sum_{k=1}^{n} \min_{1 \leq j \leq n} \hat{f}_j^2 u_k u_k^{\mathrm{T}} \right) x \right]^{\frac{1}{2}} =$$

$$= \min_{1 \leq k \leq n} \hat{f}_k \left[x^{\mathrm{T}} \underbrace{\left(\sum_{k=1}^{n} u_k u_k^{\mathrm{T}} \right)}_{\mathbf{I}_n \text{, by Prop. (7.0.2)}} x \right]^{\frac{1}{2}} = \left(\min_{1 \leq k \leq n} \hat{f}_k \right) \sqrt{\langle x, x \rangle} =$$

$$= \left(\min_{1 \leq k \leq n} \hat{f}_k \right) \|x\|_2 \Rightarrow \left(\min_{1 \leq k \leq n} \hat{f}_k \right) \|x\|_2 \leq \|\mathbf{K}_f x\|_2.$$

(b)

$$\|\mathbf{K}_f x\|_2 = \sqrt{\langle \mathbf{K}_f x, \mathbf{K}_f x \rangle} = \sqrt{x^{\mathrm{T}} \mathbf{U} \mathbf{M}_{\hat{f}}^2 \mathbf{U}^{\mathrm{T}} x} =$$

$$= \left[x^{\mathrm{T}} \left(\sum_{k=1}^{n} \hat{f}_k^2 u_k u_k^{\mathrm{T}} \right) x \right]^{\frac{1}{2}} \leq \left[x^{\mathrm{T}} \left(\sum_{k=1}^{n} \max_{1 \leq j \leq n} \hat{f}_j^2 u_k u_k^{\mathrm{T}} \right) x \right]^{\frac{1}{2}} =$$

$$= \max_{1 \leq k \leq n} \hat{f}_k \left[x^{\mathrm{T}} \underbrace{\left(\sum_{k=1}^{n} u_k u_k^{\mathrm{T}} \right)}_{\mathbf{I}_n \text{, by Prop. (7.0.2)}} x \right]^{\frac{1}{2}} = \left(\max_{1 \leq k \leq n} \hat{f}_k \right) \sqrt{\langle x, x \rangle} =$$

$$= \left(\max_{1 \leq k \leq n} \hat{f}_k \right) \|x\|_2 \Rightarrow \|\mathbf{K}_f x\|_2 \leq \left(\max_{1 \leq k \leq n} \hat{f}_k \right) \|x\|_2.$$

Then, the proof is complete. □

For a window function $f \in \mathcal{L}(G)$, the windowed Fourier transform $\mathbf{F}_f x$ of a signal x is defined in the domain $G \times \hat{G}$ as (cf. [27], [38])

$$\mathbf{F}_f x(v_i, u_k) := \sqrt{n} x^{\mathrm{T}} (\mathbf{M}_{u_k} \mathbf{C}_{e_i} f).$$

We have previously considered the convolution operator $\mathbf{C}_{e_i} f$ acting on the function f as a generalized shift of f on G. Similarly, the multiplication operator $\sqrt{n}\mathbf{M}_{u_k}$ in the windowed Fourier transform acts as a generalized modulation with respect to the Fourier basis u_k. The space-frequency analysis associated with the windowed Fourier transform employs the coefficients $\mathbf{F}_f x(v_i, u_k)$ to decompose the signal x. It has been shown in [27] that the system $\{\sqrt{n}\mathbf{M}_{u_k} C_{e_i} f \mid i, k \in \{1, \ldots, n\}\}$ provides a frame for the space of signals $\mathcal{L}(G)$ if $\hat{f}_1 \neq 0$. Assuming that the window function f is positive definite, we can refine this statement. Under this condition, the shifts $\{\mathbf{C}_{e_i} f \mid i \in \{1, \ldots, n\}\}$ constitute already a basis of $\mathcal{L}(G)$. Additionally, we obtain the following result:

Theorem 7.0.1.
Let $f \in \mathcal{P}_+$ and assume that u_1 is fixed as $u_1 = \frac{1}{\sqrt{n}}(1,\dots,1)^T$.

(1) if u_k is non-vanishing for a fixed $k \in \{1,\dots,n\}$, then the system $\{\sqrt{n}\boldsymbol{M}_{u_k} C_{e_i} f \mid i \in \{1,\dots,n\}\}$ is a basis of $\mathcal{L}(G)$.

(2) Let $\{u_{k_1},\dots,u_{k_M}\}$ be a subset of $\hat{G}$ containing $M \le n$ Fourier basis functions and $u_{k_1} = u_1$. Then $\{\sqrt{n}\boldsymbol{M}_{u_k} C_{e_i} f \mid i \in \{1,\dots,n\}, j \in \{1,\dots,M\}\}$ is a frame for G with the frame bounds

$$\left(\min_{1\le k\le n} \hat{f}_k\right)^2 \|x\|_2^2 \le \sum_{i=1}^{n}\sum_{j=1}^{M} (x^T(\sqrt{n}M_{u_{k_j}} C_{e_i} f))^2 \le \left(\sqrt{n}\max_{1\le k\le n} \hat{f}_k\right)^2 \|x\|_2^2$$

Proof:
(1) Given that $f \in \mathcal{P}_+$ and $u_k(v) \neq 0$ for all $v \in V$, it follows that both matrices $\mathbf{K}_f$ and $\mathbf{M}_{u_k}$ are invertible. Therefore, $\sqrt{n}\mathbf{M}_{u_k}\mathbf{K}_f$ is invertible and the n columns $\sqrt{n}\mathbf{M}_{u_k}\mathbf{C}_{e_i} f, i \in \{1,\dots,n\}$, constitutes a basis of $\mathcal{L}(G)$.
(2) By Proposition (7.0.3) we know that

$$\left(\min_{1\le k\le n} \hat{f}_k\right)\|x\|_2 \le \|\mathbf{K}_f x\|_2 \le \left(\max_{1\le k\le n} \hat{f}_k\right)\|x\|_2,$$

holds true for all $x \in \mathbb{R}^n$. Now, if we use $u_{k_1} = u_1 = (1,\dots,1)^{\mathrm{T}}/\sqrt{n}$, we obtain the lower bound

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{M}(x^{\mathrm{T}}(\sqrt{n}\mathbf{M}_{u_{k_j}}\mathbf{C}_{e_i} f))^2 &\ge \sum_{i=1}^{n}(x^{\mathrm{T}}(\sqrt{n}\mathbf{M}_{u_1}\mathbf{C}_{e_i} f))^2 = \\
&= \sum_{i=1}^{n}(x^{\mathrm{T}}(\sqrt{n}\frac{1}{\sqrt{n}}\underbrace{\mathrm{diag}(1,\dots,1)}_{\mathbf{I}_n}\mathbf{C}_{e_i} f))^2 = \sum_{i=1}^{n}(x^{\mathrm{T}}\mathbf{C}_f e_i)^2 = \\
&= \sum_{i=1}^{n}(x^{\mathrm{T}}\mathbf{C}_f)_i^2 = (x^{\mathrm{T}}\mathbf{C}_f)^{\mathrm{T}}(x^{\mathrm{T}}\mathbf{C}_f) = \langle x^{\mathrm{T}}\mathbf{C}_f, x^{\mathrm{T}}\mathbf{C}_f\rangle = \\
&= \|x^{\mathrm{T}}\mathbf{C}_f\|_2^2 = \|\mathbf{C}_f x\|_2^2 = \|\mathbf{K}_f x\|_2^2 \ge \\
&\ge \left(\min_{1\le k\le n}\hat{f}_k\right)^2\|x\|_2^2 \Rightarrow \left(\min_{1\le k\le n}\hat{f}_k\right)^2\|x\|_2^2 \le \sum_{i=1}^{n}\sum_{j=1}^{M}(x^{\mathrm{T}}(\sqrt{n}\mathbf{M}_{u_{k_j}}\mathbf{C}_{e_i} f))^2.
\end{aligned}$$

On the other hand, we obtain an upper bound as follows

$$\sum_{i=1}^{n}\sum_{j=1}^{M}(x^{\mathrm{T}}(\sqrt{n}\mathbf{M}_{u_{k_j}}\mathbf{C}_{e_i}f))^2 = n\sum_{j=1}^{M}\sum_{i=1}^{n}(x^{\mathrm{T}}(\mathbf{M}_{u_{k_j}}\mathbf{C}_{e_i}f))^2 = \dots$$

Let's take a closer look on the term $\sum_{i=1}^{n}(x^{\mathrm{T}}(\mathbf{M}_{u_{k_j}}\mathbf{C}_{e_i}f))^2$

$$\begin{aligned}\sum_{i=1}^{n}(x^{\mathrm{T}}(\mathbf{M}_{u_{k_j}}\mathbf{C}_{e_i}f))^2 &= \sum_{i=1}^{n}((\mathbf{M}_{u_{k_j}}x)^{\mathrm{T}}\mathbf{C}_{e_i}f))^2 = \sum_{i=1}^{n}((\mathbf{M}_{u_{k_j}}x)^{\mathrm{T}}\mathbf{C}_{f}e_i))^2 = \\ &= \sum_{i=1}^{n}((\mathbf{M}_{u_{k_j}}x)^{\mathrm{T}}\mathbf{C}_{f})_i^2 = ((\mathbf{M}_{u_{k_j}}x)^{\mathrm{T}}\mathbf{C}_{f})((\mathbf{M}_{u_{k_j}}x)^{\mathrm{T}}\mathbf{C}_{f})^{\mathrm{T}} = \\ &= (\mathbf{C}_f(\mathbf{M}_{u_{k_j}}x))^{\mathrm{T}}(\mathbf{C}_f(\mathbf{M}_{u_{k_j}}x)) = \langle \mathbf{C}_f\mathbf{M}_{u_{k_j}}x, \mathbf{C}_f\mathbf{M}_{u_{k_j}}x\rangle = \|\mathbf{C}_f\mathbf{M}_{u_{k_j}}x\|_2^2.\end{aligned}$$

Then, we get

$$\begin{aligned}\dots = n\sum_{j=1}^{M}\|\mathbf{C}_f\mathbf{M}_{u_{k_j}}x\|_2^2 &= n\sum_{j=1}^{M}\|\mathbf{K}_f\mathbf{M}_{u_{k_j}}x\|_2^2 \le \\ \le n\|\mathbf{K}_f\|_2^2\sum_{j=1}^{M}\|\mathbf{M}_{u_{k_j}}x\|_2^2 &\le n\left(\max_{1\le k\le n}\hat{f}_k\right)^2\sum_{j=1}^{M}\|\mathbf{M}_{u_{k_j}}x\|_2^2 \le \dots\end{aligned}$$

Now, we examine the term $\|\mathbf{M}_{u_{k_j}}x\|_2^2$ more closely

$$\begin{aligned}\|\mathbf{M}_{u_{k_j}}x\|_2^2 = \langle \mathbf{M}_{u_{k_j}}x, \mathbf{M}_{u_{k_j}}x\rangle &= x^{\mathrm{T}}\mathbf{M}_{u_{k_j}}^2x = \\ &= \sum_{i=1}^{n}u_{k_j}^2(i)x_i^2 = \sum_{i=1}^{n}(u_{k_j}(i)x_i)^2 \le \\ &\le \left(\sum_{i=1}^{n}u_{k_j}(i)x_i\right)^2 = (u_{k_j}^{\mathrm{T}}x)^2.\end{aligned}$$

Thus, we obtain

$$\dots \le n\left(\max_{1\le k\le n}\hat{f}_k\right)^2\sum_{j=1}^{M}(u_{k_j}^{\mathrm{T}}x)^2 = n\left(\max_{1\le k\le n}\hat{f}_k\right)^2\sum_{j=1}^{M}(\mathbf{U}_k^{\mathrm{T}}x)_j^2 =$$

$$= n \left(\max_{1 \le k \le n} \hat{f}_k \right)^2 (\mathbf{U}_k^{\mathrm{T}} x)^{\mathrm{T}} (\mathbf{U}_k^{\mathrm{T}} x) = n \left(\max_{1 \le k \le n} \hat{f}_k \right)^2 x^{\mathrm{T}} \mathbf{U}_k \mathbf{U}_k^{\mathrm{T}} x =$$

$$= n \left(\max_{1 \le k \le n} \hat{f}_k \right)^2 x^{\mathrm{T}} x = \left(\sqrt{n} \max_{1 \le k \le n} \hat{f}_k \right)^2 \|x\|_2^2 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{M} (x^{\mathrm{T}} (\sqrt{n} \mathbf{M}_{u_{k_j}} \mathbf{C}_{e_i} f))^2 \le \left(\sqrt{n} \max_{1 \le k \le n} \hat{f}_k \right)^2 \|x\|_2^2.$$

The proof is then complete. □

Note: if $M = n$, a slight modification of the previous proof yields the enhanced lower frame bound $\left(\sqrt{n} \min_{1 \le k \le n} \hat{f}_k\right)^2$. The upper and lower bounds $\left(\sqrt{n} \max_{1 \le k \le n} \hat{f}_k\right)^2$ and $\left(\sqrt{n} \min_{1 \le k \le n} \hat{f}_k\right)^2$ are optimal in this scenario. This was highlighted in [27] with an equivalent representation of the bounds.

8

Stability And Error Estimates

The purpose of this section is to establish upper bounds for both the interpolation error $|x(v) - \mathbf{I}_W x(v)|$ and the numerical condition of the interpolation. As previously stated, $\mathbf{I}_W x$ denotes the uniquely determined interpolant of a signal $x \in \mathcal{L}(G)$ in the native space $\mathcal{N}_{W,K_f}$. The approximation space $\mathcal{N}_{W,K_f}$ is constructed based on the set $W = \{w_1, \ldots, w_N\} \subset V$ of interpolation nodes and a positive definite function $f \in \mathcal{P}_+$ that defines the kernel K_f and the positive definite interpolation matrix $\mathbf{K}_{f,W}$.

8.1 Norming sets

To evaluate the accuracy of the interpolant $\mathbf{I}_W x$, it's essential to assess how effectively functions within specific subsets of $\mathcal{L}(G)$ can be reconstructed from samples obtained from the set of nodes W. This assessment is facilitated by the concept of a norming set. Within the framework of $\mathcal{L}(G)$, we examine auxiliary subspaces represented by the bandlimited signals $\mathcal{B}_M$ on G. Additionally, we introduce two projection operators $\mathbf{S}_W$ and $\mathbf{B}_M$ on $\mathcal{L}(G)$ to assist in this evaluation.

$$\mathbf{S}_W x(v) = \begin{cases} x(v) & \text{if } v \in W, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \mathbf{B}_M x = \sum_{k=1}^{M} (u_k^{\mathsf{T}} x) u_k.$$

The mapping $\mathbf{S}_W$ projects a signal x onto the subset $W \subset V$, whereas $\mathbf{B}_M$ describes the projection onto the space $\mathcal{B}_M$ of bandlimited functions.

Definition 8.1.1.
We call a subset $W = \{w_1, \ldots, w_N\}$ of V a norming set for the subspace $\mathcal{B}_M \subset \mathcal{L}(G)$ if the operator $\boldsymbol{S}_W \boldsymbol{B}_M$ is injective on the subspace $\mathcal{B}_M$.

The injectivity of $\mathbf{S}_W \mathbf{B}_M$ on $\mathcal{B}_M$ guarantees that we have a well-defined inverse $(\mathbf{S}_W \mathbf{B}_M)^{-1}$ on the image $\mathbf{S}_W(\mathcal{B}_M) = \mathbf{S}_W \mathbf{B}_M(\mathcal{L}(G))$. The operator norm of this inverse

$$\|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|_2 = \sup_{z \in \mathbf{S}_W(\mathcal{B}_M), \|z\| \leq 1} \|(\mathbf{S}_W \mathbf{B}_M)^{-1} z\|_2$$

is denoted as the norming constant for the set W. Definition (8.1.1) formally defines the following statement equivalently: W is a norming set for $\mathcal{B}_M$ if each bandlimited signal $x \in \mathcal{B}_M$ can be uniquely reconstructed from the samples $\{x(w_1), \ldots, x(w_N)\}$. In the literature, the notion "uniqueness set" is interchangeably employed instead of "norming set", as referenced in [7],[39].
A straightforward criterion outlined in [8] determines whether W qualifies as a norming set for $\mathcal{B}_M$.

Theorem 8.1.1.
The set W is a norming set for the space $\mathcal{B}_M$ of bandlimited functions if and only if the spectral norm of the matrix $\boldsymbol{B}_M(\boldsymbol{I}_n - \boldsymbol{S}_W)\boldsymbol{B}_M$ is strictly less than 1. The norming constant of the set W is bounded by

$$\|(\boldsymbol{S}_W \boldsymbol{B}_M)^{-1}\|_2 \leq \frac{1}{1 - \|\boldsymbol{B}_M(\boldsymbol{I}_n - \boldsymbol{S}_W)\boldsymbol{B}_M\|_2}$$

Proof:
Since $\mathbf{S}_W$ and $\mathbf{B}_M$ are both projection operators we know that: $\mathbf{S}_W = \mathbf{S}_W^{\mathrm{T}} = \mathbf{S}_W^2$ and $\mathbf{B}_M = \mathbf{B}_M^{\mathrm{T}} = \mathbf{B}_M^2$. The operator $\mathbf{S}_W \mathbf{B}_M$ restricted to the space $\mathcal{B}_M$ is injective if and only if

$$(\mathbf{S}_W \mathbf{B}_M)^{\mathrm{T}} \mathbf{S}_W \mathbf{B}_M = \mathbf{B}_M^{\mathrm{T}} \mathbf{S}_W^{\mathrm{T}} \mathbf{S}_W \mathbf{B}_M = \mathbf{B}_M \mathbf{S}_W^2 \mathbf{B}_M = \mathbf{B}_M \mathbf{S}_W \mathbf{B}_M$$

is invertible on $\mathcal{B}_M$(we denote its pseudo-inverse by $(\mathbf{B}_M \mathbf{S}_W \mathbf{B}_M)^{\dagger}$). This is true if and only if the extended operator $\mathbf{I}_n - \mathbf{B}_M + \mathbf{B}_M \mathbf{S}_W \mathbf{B}_M$ is invertible on the entire space $\mathcal{L}(G)$. On the other hand, this is equivalent to the fact that the spectral norm of the operator

$$\mathbf{B}_M - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M = \mathbf{B}_M^2 - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M = \mathbf{B}_M(\mathbf{I}_n - \mathbf{S}_W)\mathbf{B}_M$$

is strictly less than 1. This confirms the first statement.
For the second statement, we write the inverse $(\mathbf{S}_W\mathbf{B}_M)^{-1}$ on the image $\mathbf{S}_W(\mathcal{B}_M)$ as

$$(\mathbf{S}_W\mathbf{B}_M)^{-1}z \overset{(*)}{=} (\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{\dagger}\mathbf{B}_M\mathbf{S}_W z \overset{(*,*)}{=} (\mathbf{I}_n - \mathbf{B}_M + \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1}\mathbf{B}_M\mathbf{S}_W z.$$

Let's define $\mathbf{C} = (\mathbf{S}_W\mathbf{B}_M)^{\mathrm{T}}\mathbf{S}_W\mathbf{B}_M = \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M$ and we show equality $(*)$. Since the pseudo-inverse is define as follows $\mathbf{C}^{\dagger} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}$ and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, we have

$$\begin{aligned}
\mathbf{C}^{\dagger} &= (\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{\dagger} = ((\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{\mathrm{T}}\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1}(\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{\mathrm{T}} = \\
&= (\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1}(\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M) = \\
&= (\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1}(\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1}(\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M) = \\
&= (\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1} = (\mathbf{S}_W\mathbf{B}_M)^{-1}\mathbf{B}_M^{-1}.
\end{aligned}$$

Thus,

$$(\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{\dagger}\mathbf{B}_M\mathbf{S}_W z = (\mathbf{S}_W\mathbf{B}_M)^{-1}\mathbf{B}_M^{-1}\mathbf{B}_M\mathbf{S}_W z = (\mathbf{S}_W\mathbf{B}_M)^{-1}\mathbf{S}_W z = (\mathbf{S}_W\mathbf{B}_M)^{-1}z.$$

Here, in the last equality, we have used the fact that $\mathbf{S}_W z = z$ due to $z \in \mathbf{Im}(\mathbf{S}_W)$.
On the other hand, the equality $(*, *)$ is maintained due to the orthogonality between the transformation $(\mathbf{I}_n - \mathbf{B}_M)$ and $\mathbf{B}_M$. Indeed, $(\mathbf{I}_n - \mathbf{B}_M)\mathbf{B}_M = \mathbf{B}_M - \mathbf{B}_M^2 = \mathbf{B}_M - \mathbf{B}_M = 0_n$. Consequently, both operators $(\mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{\dagger}$ and $(\mathbf{I}_n - \mathbf{B}_M + \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^{-1}$ performs equivalently on $\mathbf{B}_M\mathbf{S}_W z$.
Now, by recalling the Neumann series expansion

$$\text{if } \|\mathbf{A}\|_2 < 1, \quad \sum_{k=0}^{n} \mathbf{A}^k = (\mathbf{I}_n - \mathbf{A})^{-1}$$

we obtain

$$(\mathbf{S}_W\mathbf{B}_M)^{-1}z = \sum_{k=0}^{\infty}(\mathbf{B}_M - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^k\mathbf{B}_M\mathbf{S}_W z.$$

Then, taking the norm on both sides gives the desired bound

$$\begin{aligned}
\|(\mathbf{S}_W\mathbf{B}_M)^{-1}z\|_2 \leq \|\sum_{k=0}^{\infty}(\mathbf{B}_M - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^k\mathbf{B}_M\mathbf{S}_W z\|_2 \leq \\
\leq \sum_{k=0}^{\infty}\|(\mathbf{B}_M - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M)^k\mathbf{B}_M\mathbf{S}_W z\|_2 \leq \\
\leq \sum_{k=0}^{\infty}\|\mathbf{B}_M - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M\|_2{}^k\|\mathbf{B}_M\mathbf{S}_W z\|_2 \leq \\
\leq \frac{1}{1-\|\mathbf{B}_M - \mathbf{B}_M\mathbf{S}_W\mathbf{B}_M\|_2}\underbrace{\|\mathbf{B}_M\|_2\|\mathbf{S}_W\|_2}_{1}\|z\|_2 = \\
= \frac{1}{1-\|\mathbf{B}_M(\mathbf{I}_n - \mathbf{S}_W)\mathbf{B}_M\|_2}\|z\|_2 \Rightarrow \\
\Rightarrow \|(\mathbf{S}_W\mathbf{B}_M)^{-1}z\|_2 \leq \frac{1}{1-\|\mathbf{B}_M(\mathbf{I}_n - \mathbf{S}_W)\mathbf{B}_M\|_2}\|z\|_2.
\end{aligned}$$

This completes the theorem. □

Theorem (8.1.1) illustrates that determining of whether W constitutes a norming set for $\mathcal{B}_M$ depends significantly on the spectral structure $\hat{G}$ of the graph G and is closely linked to the existence of non-admissible regions within the combined space-frequency domain of the graph. In spectral graph theory, these regions characterize uncertainty principles. To delve into this connection with uncertainty principles and concrete examples, consider [8] and the more general framework in [39].

8.2 Main error estimate

Our main error estimate is formulated as follows:

Theorem 8.2.1.
Let $f \in \mathcal{P}_+$ and $W \subset V$ be a norming set for the space $\mathcal{B}_M$ on the graph G. Then, for the GBF

interpolant $\boldsymbol{I}_W x \in \mathcal{N}_{W,K_f}$ of a graph signal x we get the uniform error bound

$$\max_{v\in V} |x(v) - \boldsymbol{I}_W x(v)| \leq (1 + \|(\boldsymbol{S}_W \boldsymbol{B}_M)^{-1}\|_2) \left(\sum_{k=M+1}^{n} \hat{f}_k \right)^{1/2} \|x\|_{K_f}.$$

The error estimate is influenced by three correlated factors: the norming constant $\|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|_2$, the tail $\sum_{k=M+1}^{n} \hat{f}_k$, and the native space norm $\|x\|_{K_f}$. These factors depend on the sampling set W, the bandwidth M, the decay rate of the Fourier coefficients $\hat{f}_k$ and the signal x. In terms of M and the decay rate of the coefficients $\hat{f}_k$, there exists a trade-off between the first two factors and the last two factors of the error estimate. Typically, we can obtain significant error estimates in Theorem (8.2.1) when the Fourier coefficients $\hat{f}_k$ exhibit rapid decay, the signal x is smooth (with a small native space norm $\|x\|_{K_f}$) and the sampling set W forms a norming set for a large space $\mathcal{B}_M$. Moreover, if the Fourier coefficients $\hat{f}_k$ have a particular decay, we can achieve the following additional refinements:

Corollary 8.2.1.
With the same assumption in Theorem (8.2.1), *we get the following bounds:*

(1) if $\hat{f}_k \leq C_1 k^{-s}, s > 1,$ then

$$\max_{v\in V} |x(v) - \boldsymbol{I}_W x(v)| \leq \sqrt{\frac{C_1}{s-1}} (1 + \|(\boldsymbol{S}_W \boldsymbol{B}_M)^{-1}\|_2) M^{-\frac{s-1}{2}} \|x\|_{K_f}$$

(2) if $\hat{f}_k \leq C_2 e^{-tk}, t > 0,$ then

$$\max_{v\in V} |x(v) - \boldsymbol{I}_W x(v)| \leq \sqrt{\frac{C_2}{1-e^{-t}}} (1 + \|(\boldsymbol{S}_W \boldsymbol{B}_M)^{-1}\|_2) e^{-\frac{t}{2}(M+1)} \|x\|_{K_f}$$

Proof:
This corollary directly follows from Theorem (8.2.1), along with the following observations:
(1)

$$\sum_{k=M+1}^{n} \hat{f}_k \leq C_1 \sum_{k=M+1}^{n} \frac{1}{k^s} \leq C_1 \int_M^\infty \frac{1}{x^s} dx =$$
$$= C_1 \lim_{\varepsilon\to\infty} \frac{x^{1-s}}{1-s}\Big|_M^\varepsilon = C_1 \lim_{\varepsilon\to\infty} \frac{\varepsilon^{1-s}}{1-s} - \frac{M^{1-s}}{1-s} = C_1 \frac{M^{1-s}}{s-1}$$

(2)

$$\sum_{k=M+1}^{n} \hat{f}_k \leq C_2 \sum_{k=M+1}^{\infty} e^{-tk} = C_2 \sum_{j=0}^{\infty} e^{-t(j+M+1)} =$$
$$= C_2 e^{-t(M+1)} \sum_{j=0}^{\infty} e^{-tj} = C_2 \frac{e^{-t(M+1)}}{1-e^{-t}}$$

8.3 Proof of Theorem (8.2.1)

The approach taken in this proof is inspired from similar error estimates found in the proofs for SBF's [40], for positive definite kernels on Riemannian manifolds [41] and on compact groups [19]. To estimate the error $|x(v) - \mathbf{I}_W x(v)|$, the initial step of the proof involves representing the interpolant $\mathbf{I}_W x \in \mathcal{N}_{K_f,W}$ in a suitable way. This representation is provided in terms of a Lagrange-type basis $\{\ell_1, \dots, \ell_N\}$ of $\mathcal{N}_{K_f,W}$ as

$$\mathbf{I}_W x(v) = \sum_{k=1}^{N} \ell_k(v) x(w_k). \tag{8.1}$$

The Lagrange basis functions ℓ_k are determined as the interpolants

$$\ell_k(v) = \mathbf{I}_W e_{j_k}(v), \quad k \in \{1, \dots, N\},$$

where the node v_{j_k} corresponds to the node $w_k \in W$. Now, utilizing the property that K_f operates as the reproducing kernel of the Hilbert space $\mathcal{N}_{K_f}$, we obtain the estimate

$$|x(v) - \mathbf{I}_W x(v)| = |x(v) - \sum_{k=1}^{N} \ell_k(v) x(w_k)| =$$
$$= \left| \langle x, K_f(\cdot, v) \rangle_{K_f} - \sum_{k=1}^{N} \ell_k(v) \langle x, K_f(\cdot, w_k) \rangle_{K_f} \right| =$$
$$= \left| \left\langle x, K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) K_f(\cdot, w_k) \right\rangle_{K_f} \right| \overset{C-S}{\leq}$$

$$\stackrel{C-S}{\leq} \|x\|_{K_f} \|K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) K_f(\cdot, w_k)\|_{K_f} =$$

$$= \|x\|_{K_f} \|K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f\|_{K_f}. \tag{8.2}$$

The norm $P_{W,K_f}(v) = \|K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f\|_{K_f}$ is commonly known as the power function within the RBF community (see, [16], [17]). It depends on the node v, the sampling nodes W and the positive definite function f, but it does not rely on the signal x. In order to conclude this proof, we aim to estimate the power function $P_{W,K_f}(v)$. To achieve this, we need two well-established auxiliary results. The first one is correlated to the power function.

Lemma 8.3.1. *([16], Theorem 11.1; [17], Theorem 11.5)*
If $f \in \mathcal{P}_+$, then $\sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f$ is the best approximation of $K_f(\cdot, v)$ in the subspace $\mathcal{N}_{K_f,W}$ with respect to the native space norm $\mathcal{N}_{K_f}$.

Proof:
The outcome arises from the orthogonality of the subspace $\mathcal{N}_{K_f,W}$ with the vector $K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f$. This stems from the identities

$$\left\langle K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f, \mathbf{C}_{e_{j_i}} f \right\rangle_{K_f} =$$

$$= \langle K_f(\cdot, v), \mathbf{C}_{e_{j_i}} f \rangle_{K_f} - \sum_{k=1}^{N} \ell_k(v) \langle \mathbf{C}_{e_{j_k}} f, \mathbf{C}_{e_{j_i}} f \rangle_{K_f} =$$

$$= \langle K_f(\cdot, v), K_f(\cdot, v_{j_i}) \rangle_{K_f} - \sum_{k=1}^{N} \ell_k(v) \langle K_f(\cdot, v_{j_k}), K_f(\cdot, v_{j_i}) \rangle_{K_f} =$$

$$= K_f(v, v_{j_i}) - \sum_{k=1}^{N} \ell_k(v) K_f(v_{j_k}, v_{j_i}) = K_f(w_i, v) - \sum_{k=1}^{N} \ell_k(v) K_f(w_i, v_{j_k}) =$$

$$= \mathbf{C}_{e_{j_i}} f(v) - \sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f(w_i) = \mathbf{C}_{e_{j_i}} f(v) - \mathbf{C}_{e_{j_i}} f(v) = 0.$$

The final equality arises from the definition of the Lagrange basis $\ell_k, k \in \{1, \dots, N\}$. This basis ensures that the function $\sum_{k=1}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f(w_i)$ interpolates $\mathbf{C}_{e_{j_i}} f$ at all nodes $w \in W$.

Since this interpolation is unique within $\mathcal{N}_{K_f,W}$, the summation $\sum_{k=1}^N \ell_k(v)\mathbf{C}_{e_{j_k}} f(w_i)$ corresponds to $\mathbf{C}_{e_{j_i}} f(v)$ on the entire node set V. □

The second auxiliary result is related to norming sets, which can be established through a functional analytic approach that leverages the Hahn-Banach Theorem. Full details are provided in ([17], Theorem 3.4).

Lemma 8.3.2. *([17], Theorem 3.4)*
Suppose $W = \{w_1, \ldots, w_N\}$ is a norming set for $\mathcal{B}_M \subset \mathcal{L}(G)$. Then for every node $v \in V$, there are coefficients $(\alpha_1(v), \ldots, \alpha_N(v)) \in \mathbb{R}^N$ such that

$$x(v) = \sum_{k=1}^{N} \alpha_k(v) x(w_k) \quad \textit{and} \quad \sum_{k=1}^{N} |\alpha_k(v)|^2 \leq \|(\boldsymbol{S}_W \boldsymbol{B}_M)^{-1}\|_2^2$$

for all signals $x \in \mathcal{B}_M$

Proof:
We define the linear functional $\tilde{x}$ on $\mathbf{S}_W\mathbf{B}_M(\mathcal{L}(G))$ by $\tilde{x}(y) = x((\mathbf{S}_W\mathbf{B}_M)^{-1}y)$. Its norm is bounded $\|\tilde{x}\| \leq \|x\|_{\mathcal{B}_M}\|(\mathbf{S}_W\mathbf{B}_M)^{-1}\|_2 = \|\mathbf{B}_M x\|_2\|(\mathbf{S}_W\mathbf{B}_M)^{-1}\|_2 \leq \|(\mathbf{S}_W\mathbf{B}_M)^{-1}\|_2$ since $\|\mathbf{B}_M x\|_2 \leq 1$. By the Hahn-Banach Theorem, $\tilde{x}$ has a norm-preserving extension $\tilde{x}_{\text{ext}}$ to $\mathbb{R}^N$. Due to the Riesz representation Theorem, every linear functional on $\mathbb{R}^N$ can be expressed by the inner product with a fixed vector. Consequently, there exists $\alpha(v) \in \mathbb{R}^N$ such that

$$\tilde{x}_{\text{ext}}(y) = \sum_{k=1}^{N} \alpha_k(v) y_k = \sum_{k=1}^{N} \alpha_k(v) x(w_k), \quad y_k = x(w_k) \text{ for } k = 1, \ldots, N$$

and

$$\|\alpha(v)\|_2^2 = \sum_{k=1}^{N} |\alpha_k(v)|^2 \leq \|x\|_{\mathcal{B}_M}^2 \|(\mathbf{S}_W\mathbf{B}_M)^{-1}\|_2^2 \leq \|(\mathbf{S}_W\mathbf{B}_M)^{-1}\|_2^2.$$

Finally, we find for an arbitrary $v \in V$, by setting $y = \mathbf{S}_W\mathbf{B}_M v$, the following

$$x(v) = x((\mathbf{S}_W\mathbf{B}_M)^{-1}y) = \tilde{x}(y) = \tilde{x}_{\text{ext}}(y) = \sum_{k=1}^{N} \alpha_k(v) y_k = \sum_{k=1}^{N} \alpha_k(v) x(w_k),$$

which finishes the proof.* □

*Proofs of the Hahn-Banach and the Riesz Representation Theorems can be found in the appendix.

Now, starting from the bound of the approximation error provided in equation (8.2), we proceed to evaluate the power function $P_{W,K_f}(v)$. Without loss of generality, let's assume that $v \notin W$ (since for $v \in W$, the power function is equal to zero). We set $w_0 = v_{j_0} = v$ and $\ell_0(v) = -1$. Employing the characterization of the native space norm as outlined in Theorem (5.0.1), we can represent the square of the power function as

$$\begin{aligned}
P^2_{W,K_f}(v) &= \left\| K_f(\cdot, v) - \sum_{k=1}^{N} \ell_k(v) K_f(\cdot, w_k) \right\|^2_{K_f} = \\
&= \left\| -\left(\ell_0(v) K_f(\cdot, w_0) + \sum_{k=1}^{N} \ell_k(v) K_f(\cdot, w_k) \right) \right\|^2_{k_f} = \\
&= \left\| \sum_{k=0}^{N} \ell_k(v) K_f(\cdot, w_k) \right\|^2_{K_f} = \left\| \sum_{k=0}^{N} \ell_k(v) \mathbf{C}_{e_{j_k}} f \right\|^2_{K_f} = \\
&= \sum_{l=1}^{n} \frac{1}{\hat{f}_l} \left(\sum_{k=0}^{N} \ell_k(v) \widehat{\mathbf{C}_{e_{j_k}} f} \right)^2_l = \sum_{l=1}^{n} \frac{1}{\hat{f}_l} \left(\sum_{k=0}^{N} \ell_k(v) (\mathbf{M}_{\hat{e}_{j_k}} \hat{f})_l \right)^2 = \\
&= \sum_{l=1}^{n} \hat{f}_l \left(\sum_{k=0}^{N} \ell_k(v) (\widehat{e_{j_k}})_l \right)^2 .
\end{aligned}$$

By using Lemma (8.3.1), the functional minimization of the square $P^2_{W,K_f}(v)$ is obtained with the coefficients $\ell_k(v)$. Consequently, we can establish an upper bound for $P^2_{W,K_f}(v)$ by substituting the coefficients $\ell_k(v)$ with the functions $\alpha_k(v)$, where $k \in \{1, \ldots, N\}$, as given in Lemma (8.3.2). Additionally, we define $\alpha_0(v) = -1$. Thus, we obtain the bound

$$\begin{aligned}
P^2_{W,K_f} &\overset{(a)}{\leq} \sum_{l=M+1}^{n} \hat{f}_l \left(\sum_{k=0}^{N} \alpha_k(v) (\widehat{e_{j_k}})_l \right)^2 \overset{(b)}{\leq} \sum_{l=M+1}^{n} \hat{f}_l \sum_{k=0}^{N} \alpha_k^2(v) \underbrace{\sum_{k=0}^{N} (\widehat{e_{j_k}})_l^2}_{\delta_{kl}} \leq \\
&\leq \sum_{l=M+1}^{n} \hat{f}_l \left(\alpha_0^2(v) + \sum_{k=1}^{N} |\alpha_k(v)|^2 \right) \leq (1 + \|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|_2^2) \sum_{l=M+1}^{n} \hat{f}_l \overset{(c)}{\leq} \\
&\overset{(c)}{\leq} (1 + \|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|_2)^2 \sum_{l=M+1}^{n} \hat{f}_l \Rightarrow \\
&\Rightarrow P_{W,K_f} \leq (1 + \|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|_2) \left(\sum_{l=M+1}^{n} \hat{f}_l \right)^{1/2} . \qquad (8.3)
\end{aligned}$$

Where in (a) we leverage the fact that the functions f are bandlimited and that $\ell_k(v) \leq \alpha_k(v)$ holds for all $k \in \{1, \ldots, N\}$. Furthermore, in (b) and (c), we employed the inequalities $\left(\sum_i a_i b_i\right)^2 \leq \sum_i a_i^2 \sum_i b_i^2$ and $(x^2 + y^2) \leq (x + y)^2$ if and only if $x, y > 0$ or $x, y < 0$. Thus, by applying the result (8.3) in (8.2), we obtain the desired inequality

$$\max_{v \in V} |x(v) - \mathbf{I}_W x(v)| \leq (1 + \|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|_2) \left(\sum_{k=M+1}^{n} \hat{f}_k \right)^{1/2} \|x\|_{K_f}.$$

This completes the proof. □

8.4 Stability

A commonly used measure to evaluate the absolute numerical condition of a linear interpolation scheme $\mathbf{I}_W x$ is defined by the operator norm $\sup_{\|x\|_2 \leq 1} \|\mathbf{I}_W x\|_2$. This measure quantifies the maximum potential amplification of errors in the sampled data resulting from the interpolation process. For this numerical condition number, we obtain:

Theorem 8.4.1.
If $f \in \mathcal{P}_+$, then the numerical condition number for GBF interpolation is bounded by

$$\sup_{\|x\|_2 \leq 1} \|\boldsymbol{I}_W x\|_2 \leq \|\boldsymbol{K}_{f,W}^{-1}\|_2 \|\boldsymbol{K}_f\|_2 \leq \frac{\max_{1 \leq k \leq n} \hat{f}_k}{\min_{1 \leq k \leq n} \hat{f}_k}.$$

Proof:
In matrix-vector notation we can write the interpolant $\mathbf{I}_W x$ compactly as

$$\begin{aligned} \mathbf{I}_W x(\cdot) = x(\cdot) = \langle K_f(\cdot, \cdot), x \rangle_{K_{f,W}} = K_f(\cdot, \cdot) \mathbf{K}_{f,W}^{-1} x^{\mathrm{T}} = \\ = \left(\mathbf{C}_{e_{j_1}} f, \ldots, \mathbf{C}_{e_{j_N}} f \right) \mathbf{K}_{f,W}^{-1} \left(x_{j_1}, \ldots, x_{j_N} \right)^{\mathrm{T}}. \end{aligned}$$

Therefore, we can bound the norm $\|\mathbf{I}_W x\|_2$ by

$$\begin{aligned} \|\mathbf{I}_W x\|_2 &= \|(\mathbf{C}_{e_{j_1}} f, \ldots, \mathbf{C}_{e_{j_N}}) f \mathbf{K}_{f,W}^{-1} \left(x_{j_1}, \ldots, x_{j_N} \right)^{\mathrm{T}} \|_2 \leq \\ &\leq \|(\mathbf{C}_{e_{j_1}} f, \ldots, \mathbf{C}_{e_{j_N}} f)\|_2 \|\mathbf{K}_{f,W}^{-1}\|_2 \|(x_{j_1}, \ldots, x_{j_N})^{\mathrm{T}}\|_2 \leq \|\mathbf{K}_f\|_2 \|\mathbf{K}_{f,W}^{-1}\|_2 \|x\|_2. \end{aligned}$$

According to Theorem (4.2.1), we know that the Fourier coefficients of f correspond to the eigenvalues of the positive definite matrix $\mathbf{K}_f$, and thus

$$\|\mathbf{K}_f\|_2 = \max_{1\leq k\leq n} \hat{f}_k \quad \text{and} \quad \|\mathbf{K}_f^{-1}\|_2 = \frac{1}{\min_{1\leq k\leq n} \hat{f}_k}.$$

Furthermore, based on the inclusion principle ([35], Theorem 4.3.15), the smallest eigenvalue of the principal submatrix $\mathbf{K}_{f,W}$ is larger than $\min_{1\leq k\leq n} \hat{f}_k$. Consequently, we have $\|\mathbf{K}_{f,W}^{-1}\|_2 \leq (\min_{1\leq k\leq n} \hat{f}_k)^{-1}$, which confirms the statement of the theorem. □

Hence, stability becomes an issue in GBF interpolation when the interpolation matrix $\mathbf{K}_{f,W}$ is badly conditioned. Choosing basis functions f in which the Fourier coefficients $\hat{f}_k$ are significantly far from zero helps to avoid bad conditioning. On the other hand, for the error bounds outlined in Theorem (8.2.1), it's crucial that the Fourier coefficients decay quickly towards zero. This illustrates the trade-off between stability and approximation quality in the method, a phenomenon typically observed in classical RBF and SBF interpolation, as discussed in [17], [33], [42].

8.5 Numerical example

To assess the effectiveness of GBF interpolation performance compared to a pure bandlimited interpolation, we present two numerical examples. We employ a test graph, denoted as G, which is a reduced point cloud obtained from the Stanford bunny model (sourced from Stanford University Computer Graphics Laboratory). This graph contains $n = 900$ nodes projected onto the xy-plane, with 7325 edges connecting them. Two nodes are connected by an edge if their Euclidean distance is smaller than a given radius of 0.01. We recursively create a sequence W_N of N sampling sets in V such that the cardinality of W_N is equals to N, $\#W_N = N$. Each set W_{N-1} is contained in W_N, and the new node w_N in W_N is chosen randomly from $V \setminus W_{N-1}$. For our first test signal, we employ the signal $x^{(1)} = u_4$ as depicted in figure (8.1), which represents a bandlimited test function in the space $\mathcal{B}_4$. As a second illustration, we use a non-bandlimited, smooth signal $x^{(2)}$ shown in figure (8.2). The Fourier coefficients $\hat{x}_k^{(2)}$ of $x^{(2)}$ decay exponentially with k.
The signal $x^{(1)} = u_4$ can be exactly reconstructed within the space $\mathcal{B}_N$ if $N \geq 4$ and W_N is a norming set for $\mathcal{B}_N$. This is evident in figure (8.1). On the other hand, as depicted in figure (8.2), interpolation in $\mathcal{B}_N$ becomes highly unstable when the signal $x^{(2)}$ is outside of $\mathcal{B}_N$, even if $x^{(2)}$ is extremely smooth. However, The GBF interpolants exhibit similarly stable behavior

in both scenarios. In addition, as predicted by Theorem (8.2.1), the outcomes in figure (8.1) and (8.2) show that the Fourier decay of the various GBF's significantly affects the convergence rates when the interpolated signals are smooth.

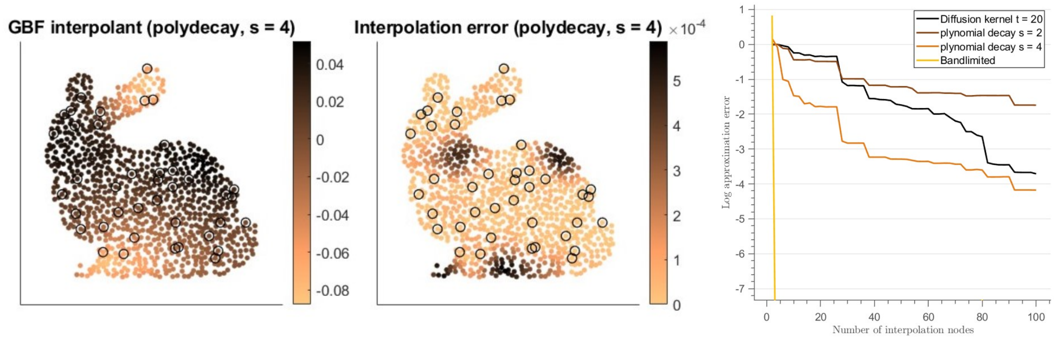


Figure 8.1: GBF interpolation for the input signal $x^{(1)} = u_4$. Left: GBF interpolant for the nodes W_{40} and the GBF $f_{\mathrm{pol},4}$ given in Example **(6)** from Chapter (6). Middle: interpolation error compared to the original signal. Right: interpolation errors for GBF schemes relative to the number N of interpolation nodes.

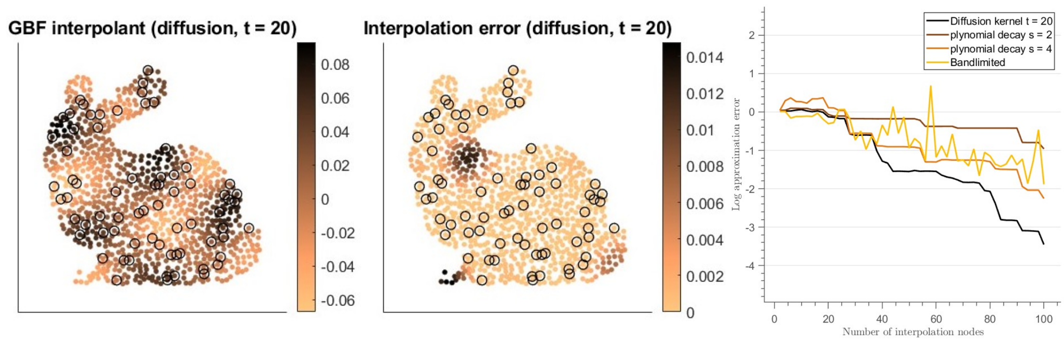


Figure 8.2: GBF interpolation for the input signal $x^{(2)}$. Left: GBF interpolant for the nodes W_{70} and the diffusion GBF $f_{e^{-20L}}$ given in Example **(5)** from Chapter (6). Middle: interpolation error with respect to the original signal. Right: Interpolation errors for GBF schemes in terms of the number N of interpolation nodes.

9

Integration Of Graph Signals With Positive Definite Functions

As a final application of positive definite functions on graphs, our goal is to determine quadrature weights μ_k, $k \in \{1,\dots,N\}$ such that the integration functional $\frac{1}{n}\sum_{i=1}^{n} x(v_i)$ is well approximated by a sum of the form $\sum_{k=1}^{N} \mu_k x(w_k)$. Here, $W = \{w_1,\dots,w_N\}$ is a subset of V. Following the approach suggested for variational splines [43], we construct the quadrature weights to ensure that the quadrature formula is exact for all signals in the interpolation space $\mathcal{N}_{K_f,W}$, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} x(v_i) = \sum_{k=1}^{N} \mu_k x(w_k) \quad \text{for all } x \in \mathcal{N}_{K_f,W}. \tag{9.1}$$

As previously mentioned, f is positive definite. GBF provides the basis $\left\{\mathbf{C}_{e_{j_1}}, \mathbf{C}_{e_{j_2}}, \dots, \mathbf{C}_{e_{j_N}}\right\}$ for the space $\mathcal{N}_{K_f,W}$. The indices j_k are determined by the relation $v_{j_k} = w_k$.
The exactness in (9.1) gives us a system of equations to find the coefficients μ_k, where $k \in \{1,\dots,N\}$. To obtain this system, we assume that the first eigenvector u_1 of the Laplacian $\mathbf{L}$ is given by $u_1 = (1,\dots,1)^{\mathrm{T}}/\sqrt{n}$. Thus, plugging the basis functions $\mathbf{C}_{e_{j_k}} f$ into equation

(9.1), we obtain the following identities

$$\frac{1}{\sqrt{n}}\underbrace{u_{j_k}(v_1)}_{\mathbf{C}_{e_{j_k}}}\hat{f}_1 = \frac{1}{\sqrt{n}}\left(\widehat{\mathbf{C}_{e_{j_k}} f}\right)_1 = \frac{1}{\sqrt{n}}(u_1^{\mathrm{T}}\mathbf{C}_{e_{j_k}} f) =$$

$$= \frac{1}{n}(1,\ldots,1)\mathbf{C}_{e_{j_k}} f = \frac{1}{n}\sum_{i=1}^{n}\mathbf{C}_{e_{j_k}} f(v_i) = \sum_{l=1}^{N}\mu_l \mathbf{C}_{e_{j_k}} f(w_l)$$

for $k \in \{1,\ldots,N\}$. Combining these N identities and using the fact that $\mathbf{C}_{e_{j_k}} f(w_l) = \mathbf{C}_{e_{j_l}}(w_k)$, we obtain the linear system of equations

$$\underbrace{\begin{pmatrix} \mathbf{C}_{e_{j_1}} f(w_1) & \cdots & \mathbf{C}_{e_{j_N}} f(w_1) \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{e_{j_1}} f(w_N) & \cdots & \mathbf{C}_{e_{j_N}} f(w_N) \end{pmatrix}}_{\mathbf{K}_{f,W}} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} = \frac{1}{\sqrt{n}}\hat{f}_1 \begin{pmatrix} u_{j_1}(v_1) \\ \vdots \\ u_{j_N}(v_1) \end{pmatrix}. \tag{9.2}$$

Given that f is positive definite, the matrix $\mathbf{K}_{f,W}$ is invertible and the coefficients $\mu_1,\ldots,\mu_N$ are uniquely determined.

Corollary 9.0.1.
Let $f \in \mathcal{P}$ and $W \subset V$ be a norming set for the space $\mathcal{B}_M$ on the graph G. Furthermore, let the quadrature rule $\boldsymbol{Q}_W x = \sum_{k=1}^{N}\mu_k x(w_k)$ be exact for all signals in $\mathcal{N}_{K_f,W}$. Then, for $z \in \mathcal{L}(G)$, we have the error bound

$$\left|\frac{1}{n}\sum_{i=1}^{n} x(v_i) - \boldsymbol{Q}_W x\right| \leq (1 + \|(\boldsymbol{S}_W\boldsymbol{B}_M)^{-1}\|)\left(\sum_{k=M+1}^{n}\hat{f}_k\right)^{1/2}\|x\|_{K_f}.$$

Proof:
Since the quadrature formula is exact for all elements of $\mathcal{N}_{K_f,W}$, we obtain the following identities for the interpolant $\mathbf{I}_W x \in \mathcal{N}_{K_f,W}$ of a signal x:

$$\mathbf{Q}_W x = \mathbf{Q}_W \mathbf{I}_W x = \sum_{k=1}^{N}\mu_k \mathbf{I}_W x(w_k) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_W x(v_i).$$

Therefore,

$$\left|\frac{1}{n}\sum_{i=1}^{n} x(v_i) - \mathbf{Q}_W x\right| = \left|\frac{1}{n}\sum_{i=1}^{n} \underbrace{(x(v_i) - \mathbf{I}_W x(v_i))}_{\leq \max_{x\in V} |x(v_i) - \mathbf{I}_W x(v_i)|}\right| \leq \max_{x\in V} |x(v_i) - \mathbf{I}_W x(v_i)|,$$

and the given bound follows from the Theorem (8.2.1)

$$\left|\frac{1}{n}\sum_{i=1}^{n} x(v_i) - \mathbf{Q}_W x\right| \leq (1 + \|(\mathbf{S}_W \mathbf{B}_M)^{-1}\|) \left(\sum_{k=M+1}^{n} \hat{f}_k\right)^{1/2} \|x\|_{K_f}.$$

This completes the proof. □

For the variational spline kernel $f_{(\varepsilon \mathbf{I}_n + \mathbf{L})^{-s}}$ discussed in Example **(4)** of Chapter (6), the bound in Corollary (9.0.1) appears to be complementary to the quadrature error provided in [43], Theorem 3.3. On the other hand, in [5], [13], [43] a Λ-set terminology is employed to characterize the interpolation and quadrature quality of variational splines. In contrast, we have used the complementary concept of norming sets to infer the bounds in Theorem (8.2.1) and Corollary (9.0.1). While, for bandlimited functions, an additional intriguing quadrature rule related to kernels based on powers of the graph Laplacian is outlined in [44].

A
Hahn-Banach Theorem

Definition A.0.1. ***(Dual Space)***
The dual space of a vector space V is the space of all linear functionals on V, usually denoted by V'. These functionals map vectors from V to the underlying scalar field (often real numbers or complex numbers) in a linear manner.

Theorem A.0.1. ***(Hahn-Banach)***
If G is a subspace of a normalized vector space E and g is a bounded linear functional on G, then g can be extended to a linear functional f on E such that $\|f\| = \|g\|$.

Before presenting the proof, we provide some comments. Firstly, to extend g to a function f (in the most general situation) means that the domain of f includes the domain of g and $f(x) = g(x)$ for every x in the domain of g. Secondly, the norms $\|g\|$ and $\|f\|$ are computed with respect to the domains of g and f, respectively; explicitly:

$$\|g\| = \sup\left\{\frac{|g(x)|}{\|x\|} \mid x \in G\right\}, \quad \|f\| = \sup\left\{\frac{|f(x)|}{\|x\|} \mid x \in E\right\}.$$

We will prove the Hahn-Banach Theorem in two steps, assuming successively:

(i) E is a real vector space, g is a linear map with real values satisfying appropriate conditions;

(ii) E is a normalized real vector space, g is a continuous linear map with real values.

The "heart" of the proof of the theorem lies in the first step **(i)**, point **(ii)** is an important corollary of **(i)**. For the proof of the theorem, we use Zorn's lemma, which we recall below after mentioning some notions from the theory of ordered sets.

- **(a)** Let $(P, \leq)$ be a set equipped with a partial order relation. We say that a subset $Q \subset P$ is totally ordered if for every $a, b \in Q$, at least one of the relations $a \leq b$ or $b \leq a$ holds.
- **(b)** Let Q be a subset of P. We say that $c \in P$ is an upper bound of Q if for every $a \in Q$, we have $a \leq c$.
- **(c)** We say that $m \in P$ is a maximal element of P if for every $x \in P$ such that $m \leq x$, we have $x = m$.
- **(d)** Finally, we say that P is inductive if every totally ordered subset of P admits an upper bound.

Lemma A.0.1. ***(Zorn)***
Every non-empty, inductive, ordered set admits a maximal element.[*]

Theorem A.0.2. ***(Hahn-Banach - analytic form)***
Let E be a vector space over $\mathbb{R}$ and let $p : E \to \mathbb{R}$ be a function satisfying the following conditions:

$$p(\lambda x) = \lambda p(x) \ \forall\, x \in E \,\forall\, \lambda > 0, \tag{A.1}$$

$$p(x + y) \leq p(x) + p(y) \ \forall\, x, y \in E. \tag{A.2}$$

Moreover, let $G \subset E$ be a vector subspace and $g : G \to \mathbb{R}$ be a linear map such that:

$$g(x) \leq p(x) \ \forall\, x \in G.$$

Then there exists a linear map $f : E \to \mathbb{R}$ that extends g, i.e.,

$$f(x) = g(x) \ \forall\, x \in G$$

and such that

$$f(x) \leq p(x) \ \forall x \in E.$$

[*] A proof of Zorn's lemma can be found starting from the axiom of choice in N. Dunford and J. Schwartz's "Linear Operators", Volume 1, Theorem 1.2.7. [45]

Proof:
A linear application h with real values defined on a subspace of E is completely determined by such subspace, which we denote as $D(h)$, and by the law $h : D(h) \to \mathbb{R}$. Let's consider the set

$$\mathcal{P} = \left\{ h : D(h) \to \mathbb{R} \text{ such that } \begin{array}{l} \text{i) } D(h) \text{ is a vector subspace of } E; \\ \text{ii) } G \subset D(h); \\ \text{iii) } h : D(h) \to \mathbb{R} \text{ is linear;} \\ \text{iv) } h \text{ extends } g; \\ \text{v) } h(x) \leq p(x), \text{ for all } x \in D(h). \end{array} \right\}.$$

Let's equip $\mathcal{P}$ with the order relation:

$$h_1 \leq h_2 \iff D(h_1) \subset D(h_2) \text{ and } h_2 \text{ extends } h_1.$$

It's clear that $\mathcal{P}$ is not empty since $g \in \mathcal{P}$. Let's prove that $\mathcal{P}$ is inductive. Let $Q = \{h_i\}_{i \in I}$, where I is an arbitrary set of indices, be a totally ordered set of elements of $\mathcal{P}$. Let's define

$$D(h) = \bigcup_{i \in I} D(h_i) \text{ and } h(x) = h_i(x) \text{ if } x \in D(h_i).$$

We observe that h is well-defined:

(i) $h \in \mathcal{P}$

(ii) If $x \in D(h_i)$ and $x \in D(h_j)$, with $h_i, h_j \in Q$, then either $h_i \leq h_j$ or $h_j \leq h_i$. In the first scenario h_j is an extension of $h_i \Rightarrow h_j(x) = h(x) \Rightarrow h(x)$ does not depend by $h_i \in Q$. While in the second hypothesis h_i is an extension of $h_j \Rightarrow h_i(x) = h(x) \Rightarrow h(x)$ does not depend by $h_j \in Q$

(iii) $D(h)$ is a vector subspace of E, h is linear and extends g, $h(x) \leq p(x) \, \forall \, x \in D(h)$

(iv) h is an upper bound of Q

By Zorn's lemma, $\mathcal{P}$ admits a maximal element, denoted by f. Let's prove that the domain of f is E; this will complete the proof. Let's reason by contradiction and suppose that $D(f) \subsetneq E$.

Let $x_0 \in E \setminus D(f)$. Let's define a map h as follows:

$$\begin{cases} D(h) = D(f) + \mathbb{R}x_0, \\ h(x + tx_0) = f(x) + t\alpha, \;\; x \in D(f), t \in \mathbb{R}, \end{cases}$$

where α is a constant to be chosen such that $h \in \mathcal{P}$. For this purpose, it should be:

$$h(x + tx_0) \leq p(x + tx_0) \;\forall\, x \in D(f),\; \forall\, t \in \mathbb{R},$$

namely

$$f(x) + t\alpha = h(x + tx_0) \leq p(x + tx_0) \;\forall\, x \in D(f),\; \forall\, t \in \mathbb{R}. \tag{A.3}$$

Thanks to hypothesis (A.1), it suffices to verify that (A.3) holds for $t = 1$ and for $t = -1$, i.e., that:

$$\begin{cases} f(x) + \alpha \leq p(x + x_0), \;\; \forall\, x \in D(f) \\ f(x) - \alpha \leq p(x - x_0), \;\; \forall\, x \in D(f) \end{cases}.$$

Indeed for $t > 0$ and $t < 0$ we have:

$$\begin{aligned} f(x) + t\alpha &= t\left(f\left(\frac{x}{t}\right) + \alpha\right) \leq tp\left(\frac{x}{t} + \alpha\right) = p(x + t\alpha), \\ f(x) - (-t)\alpha &= -t(f\left(-\frac{x}{t}\right) - \alpha) \leq -tp\left(-\frac{x}{t} - \alpha\right) = p(x + t\alpha.) \end{aligned}$$

It remains to be shown that it is possible to choose α in such a way that

$$\begin{aligned} f(y) - p(y - x_0) \leq \alpha \;\forall\, y \in D(f), \\ \alpha \leq p(x + x_0) - f(x) \;\forall\, x \in D(f). \end{aligned}$$

Therefore

$$\sup_{y \in D(f)} \{f(y) - p(y - x_0)\} \leq \alpha \leq \inf_{x \in D(f)} \{p(x + x_0) - f(x)\}.$$

Such a choice is feasible since it holds

$$f(y) - p(y - x_0) \leq p(x + x_0) - f(x) \ \forall\, x \in D(f),\ \forall\, y \in D(f),$$

in fact, thanks to (A.2), it follows:

$$f(x) + f(y) = f(x + y) \leq p(x + y) \leq p(x) + p(y) \leq p(x + x_0) + p(y - x_0). \quad \text{(A.4)}$$

This concludes that f is bounded above by h and that $f \neq h$; this contradicts the maximality of f. □

Corollary A.0.1.
Let G a vector subsapce of E and $g : G \to \mathbb{R}$ be a continous linear function with the following norm:

$$\|g\|_{G'} = \sup_{x \in G,\ \|x\|_E \leq 1} |g(x)|.$$

Then there exists $f \subset E'$ that extends g such that

$$\|f\|_{E'} = \|g\|_{G'}.$$

Proof:
We applied the Hahn-Banach Theorem (A.0.2) with $p(x) = \|g\|_{G'}\|x\|_E \ \forall\, x \in E$, then $g(x) \leq \|g\|_{G'}\|x\|_E$. By Hahn-Banach Theorem there exists a linear functional $f : E \to \mathbb{R}$ such that $f_{|G} = g$ and $f(x) \leq p(x)\, \forall\, x \in E$. Therefore, $f(x) \leq \|g\|_{G'}\|x\|_E$.

$$\begin{aligned} &\forall\, x \in E \Rightarrow f(-x) \leq \|g\|_{G'}\|{-x}\|_E \Rightarrow |f(x)| \leq \|g\|_{G'}\|x\|_E \Rightarrow \\ &\Rightarrow \begin{cases} f \in E', \\ \|f\|_{E'} \leq \|g\|_{G'}, \text{ since } \|x\|_{E'} \leq 1 \end{cases}. \end{aligned}$$

Clearly, $\|f\|_{E'} \geq \|g\|_{G'} \Rightarrow \|f\|_{E'} = \|g\|_{G'}$. □

B

Riesz Representation Theorem

Theorem B.0.1. ***(Riesz Representation)***
Let $\varphi \in H'$ be a linear and continuous functional on the Hilbert space H. Then there exists a unique $v \in H$ such that $\varphi(x) = \langle x, v \rangle$ for every $x \in H$. Furthermore we have $\|\varphi\|_{H'} = \|v\|_H$

Proof:
Let's see the uniqueness. if $\langle x, v_1 \rangle = \langle x, v_2 \rangle$ for every $x \in H$, then the difference $u := v_1 - v_2$ is orthogonal to the entire space H. In particular, we observe that $v_1 - v_2$ is orthogonal to itself, which implies that it must be the zero vector. Thus, $v_1 = v_2$.
Let's see the existence. if φ it is identically zero, then we simply take $v = 0$. if φ it is not identically zero, then there exists a vector $u \in (\ker(\varphi))^{\perp}$ such that $\varphi(u) = 1$. For any arbitrary $x \in H$, by linearity, we have

$$\varphi(x - \varphi(x)u) = \varphi(x) - \varphi(x)\varphi(u) = \varphi(x) - \varphi(x) = 0.$$

Therefore $x - \varphi(x)u \in \ker(\varphi)$. Since u is orthogonal to the kernel $\ker(\varphi)$, we have

$$0 = \langle x - \varphi(x)u, u \rangle = \langle x, u \rangle = \varphi(x)\|u\|_2^2,$$

from which we infer that $\varphi(x) = \|u\|_2^{-2}\langle x, u \rangle$. Therefore, we obtain $\varphi(x) = \langle x, v \rangle$ for all $x \in H$, with $v = \|u\|_2^{-2}u$.

Now, to establish the equality $\|\varphi\|_{H'} = \|v\|_H$, let's define the following

$$\|\varphi\|_{H'} = \sup_{x \in H,\, x \neq 0} \frac{|\varphi(x)|}{\|x\|_H},$$
$$\|x\|_H = \sup_{\varphi \in H',\, \varphi \neq 0} \frac{|\varphi(x)|}{\|\varphi\|_{H'}}.$$

To demonstrate the equality, we show **(a)** $\|v\|_H \leq \|\varphi\|_{H'}$ and **(b)** $\|v\|_H \geq \|\varphi\|_{H'}$.

(a) $\|\varphi\|_{H'} \geq |\varphi(x)\| / \|x\|_H = |\langle x, v\rangle\| / \|x\|_H$ for all $x \in H$. Given that $v \in H$, we can choose $x = v$. Then, we have that $\|\varphi\|_{H'} \geq \|v\|_H^2 / \|v\|_H = \|v\|_H \Rightarrow \|\varphi\|_{H'} \geq \|v\|_H$.

(b) By Cauchy-Schwarz inequality, we have the following: $|\varphi(x)| = |\langle x, v\rangle| \leq \|x\|_H \|v\|_H$ for all $x \in H \Rightarrow \|v\|_H \geq |\varphi(x)| / \|x\|_H$ for all $x \in H, x \neq 0 \Rightarrow$ since the inequality holds true for all $x \in H$ with $x \neq 0$, we have that $\|v\|_H \geq \sup_{\substack{x \in H \\ x \neq 0}} \frac{|\varphi(x)|}{\|x\|_H} = \|\varphi\|_{H'} \Rightarrow \|v\|_H \geq \|\varphi\|_{H'}$.

This completes the proof. $\square$

References

[1] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[2] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[3] L. Stanković, M. Daković, and E. Sejdić, "Introduction to graph signal processing," *Vertex-Frequency Analysis of Graph Signals*, pp. 3–108, 2019.

[4] M. Belkin and P. Niyogi, "Semi-supervised learning on riemannian manifolds," *Machine learning*, vol. 56, pp. 209–239, 2004.

[5] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.

[6] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE transactions on signal processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[7] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5445–5449, IEEE, 2013.

[8] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[9] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples.," *Journal of machine learning research*, vol. 7, no. 11, 2006.

[10] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proceedings of the 19th international conference on machine learning*, vol. 2002, pp. 315–322, 2002.

[11] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 764–778, 2016.

[12] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003. Proceedings*, pp. 144–158, Springer, 2003.

[13] I. Pesenson, "Variational splines and paley–wiener spaces on combinatorial graphs," *Constructive Approximation*, vol. 29, pp. 1–21, 2009.

[14] J. P. Ward, F. J. Narcowich, and J. D. Ward, "Interpolating splines on graphs for data science applications," *Applied and Computational Harmonic Analysis*, vol. 49, no. 2, pp. 540–557, 2020.

[15] M. D. Buhmann, "Radial basis functions," *Acta numerica*, vol. 9, pp. 1–38, 2000.

[16] R. Schaback, "Native hilbert spaces for radial basis functions i," in *New Developments in Approximation Theory: 2nd International Dortmund Meeting (IDoMAT)'98, Germany, February 23–27, 1998*, pp. 255–282, Springer, 1999.

[17] H. Wendland, *Scattered data approximation*. Cambridge university press, 2005.

[18] I. J. Schoenberg, "Positive definite functions on spheres," 1942.

[19] W. Erb and F. Filbir, "Approximation by positive definite functions on compact groups," *Numerical Functional Analysis and Optimization*, vol. 29, no. 9-10, pp. 1082–1107, 2019.

[20] S. Hubbert, Q. T. Lê Gia, T. M. Morton, *et al.*, *Spherical radial basis functions, theory and applications*. Springer, 2015.

[21] H. Mhaskar, F. Narcowich, J. Prestin, and J. Ward, "$\mathcal{L}^p$-bernstein estimates and approximation by spherical basis functions," *Mathematics of Computation*, vol. 79, no. 271, pp. 1647–1679, 2010.

[22] Z. Sasvári, "Positive definite and definitizable functions," *Akademie Verlag, Berlin*, 1994.

[23] J. Stewart, "Positive definite functions and generalizations, an historical survey," *The Rocky Mountain Journal of Mathematics*, vol. 6, no. 3, pp. 409–434, 1976.

[24] S. Bochner, "Monotone funktionen, stieltjessche integrale und harmonische analyse," *Mathematische Annalen*, vol. 108, no. 1, pp. 378–410, 1933.

[25] W. Erb, "Graph signal interpolation with positive definite graph basis functions," *Applied and Computational Harmonic Analysis*, vol. 60, pp. 368–395, 2022.

[26] F. R. Chung, *Spectral graph theory*, vol. 92. American Mathematical Soc., 1997.

[27] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 260–291, 2016.

[28] K. R. Davidson, *C*-algebras by example*, vol. 6. American Mathematical Soc., 1996.

[29] B. Li, *Real operator algebras*. Fields Institute Monographs, World Scientific Publishing, Singapore, 2003.

[30] M. Puschel and J. M. Moura, "Algebraic signal processing theory: Foundation and 1-d time," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3572–3585, 2008.

[31] R. Schaback and H. Wendland, "Algebraic signal processing theory: Foundation and 1-d time," *Advanced Problems in Constructive Approximation, Birkhäuser Verlag, Basel*, pp. 203–222, 2003.

[32] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[33] G. E. Fasshauer, "Positive definite kernels: past, present and future," *Dolomites Research Notes on Approximation*, vol. 4, pp. 21–63, 2011.

[34] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American mathematical society*, vol. 68, no. 3, pp. 337–404, 1950.

[35] H. RA and J. CR, "Matrix analysis," 1985.

[36] G. Berschneider and W. zu Castell, "Conditionally positive definite kernels and pontryagin spaces," in *Approximation theory XII: San Antonio 2007*, pp. 27–37, Citeseer, 2008.

[37] T. Hangelbroek, F. J. Narcowich, and J. D. Ward, "Polyharmonic and related kernels on manifolds: interpolation and approximation," *Foundations of Computational Mathematics*, vol. 12, pp. 625–670, 2012.

[38] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "A windowed graph fourier transform," *IEEE Statistical Signal Processing Workshop (SSP)*, pp. 133–136, 2012.

[39] W. Erb, "Shapes of uncertainty in spectral graph theory," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 1291–1307, 2020.

[40] K. Jetter, J. Stöckler, and J. Ward, "Error estimates for scattered data interpolation on spheres," *Mathematics of Computation*, vol. 68, no. 226, pp. 733–747, 1999.

[41] N. Dyn, F. J. Narcowich, and J. D. Ward, "Variational principles and sobolev-type estimates for generalized interpolation on a riemannian manifold," *Constructive approximation*, vol. 15, pp. 175–208, 1999.

[42] S. De Marchi and R. Schaback, "Stability of kernel-based interpolation," *Advances in Computational Mathematics*, vol. 32, pp. 155–161, 2010.

[43] I. Z. Pesenson, M. Z. Pesenson, *et al.*, "Cubature formulas on combinatorial graphs," *arXiv preprint arXiv:1104.0963*, 2011.

[44] G. Linderman and S. Steinerberger, "Numerical integration on graphs: where to sample and how to weigh," *Mathematics of computation*, vol. 89, no. 324, pp. 1933–1952, 2020.

[45] N. Dunford and J. T. Schwartz, *Linear operators, part 1: general theory*, vol. 10. John Wiley & Sons, 1988.