Interpolating splines on graphs for data science applications

J. P. Ward†   F. J. Narcowich†   J. D. Ward‡

October 15, 2018

Abstract

We introduce intrinsic interpolatory bases for data structured on graphs and derive properties of those bases. Polyharmonic Lagrange functions are shown to satisfy exponential decay away from their centers. The decay depends on the density of the zeros of the Lagrange function, showing that they scale with the density of the data. These results indicate that Lagrange-type bases are ideal building blocks for analyzing data on graphs, and we illustrate their use in kernel-based machine learning applications.

1 Introduction

Graph, or network, domains are being used for many signal processing applications [16]. They provide a more general framework than integer lattices, and they can be used to incorporate additional structural or geometric information.

Our purpose here is to develop and analyze intrinsic interpolatory and near-interpolatory bases for graphs, with the goal of introducing more approximation tools for graphs. Kernels on graphs were introduced in [17] within the context of regularization operators. Both Gaussian and regularized Laplacian kernels were considered. Then, interpolants on graphs were considered in [14], where the author defines variational splines and Lagrangian splines that are similar to the interpolants that we propose. There, a perturbation factor is used to modify the Laplacian into a positive definite matrix, making the basis functions analogs of the Matérn functions. Our approach is to work directly with the Laplacian, making our splines analogs of the polyharmonic splines. Another distinction is that we work on weighted graphs. Properties of splines are also discussed in [15]. In particular, the author considers approximation in Paley-Wiener spaces. Note that while we are considering Lagrange functions, our approach (based on radial basis function theory) is different than that of [14, 15, 13], and hence our notation is also different.

Our main result is an estimate on the decay of Lagrange functions as this is a key first step toward related computationally efficient bases on graphs. Our reference point is the theory of Lagrange bases on Euclidean domains and Riemannian manifolds [8, 9, 19]. On \( \mathbb{R}^d \), for example, it is known that polyharmonic interpolating splines satisfy fast decay rates and form stable bases of \( L_p \) spaces, among other desirable properties of approximants. Moreover, on those continuous domains, it is known that more computationally efficient perturbations of the Lagrange functions

---

*This research was supported by grants DMS-1514789 and DMS-1813091 from the National Science Foundation.
†North Carolina A&T State University, Greensboro, NC.
‡Texas A&M University, College Station, TX.
(local Lagrange functions) satisfy similar properties. We believe all of these results should carry over to the graph setting.

While we use the continuous domain theory as a theoretical guideline, we expect the continuous and discrete theories to be complementary. For example, there are empirically estimated constants in the continuous domain results that should be more accessible for graph domains. One such constant determines how many basis functions are needed to construct a good local Lagrange function. If these constants could be worked out precisely for graphs, it could shed light on the situation in other settings.

Unlike the continuous domain, polyharmonic functions on graphs are not only bounded but also exhibit fast decay [18], so the decay of the Lagrange functions could be deduced from this. Our goal is to obtain more precise estimates of decay that depend on the density of the data, showing that the Lagrange functions are actually much more localized than the original basis functions.

There are numerous potential applications of interpolants and quasi-interpolants on graphs. In section 5 we discuss one of these, namely kernel-based machine learning. Classification problems based on many parameters seem to be a natural application for a graph model. The space of parameters likely does not live in a natural continuous ambient space. Using graphs, we can define the relationships that make sense and specify precisely how connected two objects (vertices) are by a weighted edge. We present an algorithm for such problems and apply it to sample data sets.

The remainder of the paper is organized as follows. We conclude this section with some notation and background information about functions on graphs. In section 2 we introduce the polyharmonic basis functions and verify some basic properties. section 3 contains our main results: the construction and properties of polyharmonic Lagrange functions on graphs. We define local Lagrange functions in section 4. In section 5, we illustrate our results with some examples, and we conclude in section 6.

1.1 Setting

The general setting is a finite, connected, weighted graph $G = \{V, E \subseteq V \times V, w, \rho\}$ where $V$ is the vertex set, $E$ is the edge set, and $w : V \times V \rightarrow \mathbb{R}_{\geq 0}$ is a weight function. The weight function $w$ specifies the adjacency matrix $A$ of the graph. The $j$th diagonal entry of the diagonal matrix $D$ is equal to the $j$th row sum of $A$. The function $\rho : V \times V \rightarrow \mathbb{R}_{\geq 0}$ is a distance function on the graph. Given the distance between adjacent vertices, the distance between non-adjacent vertices is the length of the shortest path connecting them.

The normalized graph Laplacian is $L = D^{-1/2}(D - A)D^{-1/2}$ [1, 2]. Let $\{\Lambda_k\}_{k=0}^{N-1}$ and $\{\lambda_k\}_{k=0}^{N-1}$ denote the eigenvectors and eigenvalues of the Laplacian. The Laplacian is a positive semi-definite matrix, with a zero eigenvalue of multiplicity one (since the graph is connected). We order the eigenvalues as

$$0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}. \quad (1)$$

The eigenvector $\Lambda_0$ is constant.

**Example 1.1.** Let $G$ be the cycle graph with $N \in \mathbb{Z}_{\geq 1}$ nodes and edges of equal weight 1. In this case the Laplacian is the circulant matrix with rows $\ldots, 0, 0, -1/2, 1, -1/2, 0, 0, \ldots$.

On a graph $G$, we denote the closed ball of radius $r$ centered at $v$ as $B(v; r)$. The annulus centered at $v$ with inner and outer radii $r_0$ and $r_1$ respectively is denoted as $A(v; r_0, r_1)$. The annulus is defined by the set difference $B(v; r_1) \setminus B(v; r_0)$.
1.2 Background: Fourier transform and smoothness spaces

A function on a graph is a mapping 
\[ f : V \rightarrow \mathbb{R}. \]  
(2)

The Lebesgue space of square-summable functions is denoted as \( \ell_2(G) \). We shall primarily focus on (semi-)Hilbert spaces on graphs, and we measure smoothness using the Laplacian. For any \( \alpha > 0 \), we define the Sobolev semi-norm
\[ |f|_{H^\alpha_2(G)} = \|L^{\alpha/2}f\|_{\ell_2(G)}. \]  
(3)

The Sobolev space \( H^\alpha_2(G) \) is defined by the norm
\[ \|f\|_{H^\alpha_2(G)} = \|f\|_{\ell_2(G)} + |f|_{H^\alpha_2(G)}. \]  
(4)

2 Conditionally positive definite basis functions

On Euclidean spaces, polyharmonic splines are defined as Green’s functions of powers of the Laplacian. On the graph, we use the column vectors of powers of \( L^\dagger \), the Moore-Penrose pseudo-inverse of the Laplacian.

**Definition 2.1.** On a graph \( G \), the order \( 2\alpha \) polyharmonic spline centered at \( v_k \in V \) is denoted as \( \Phi_\alpha(\cdot, v_k) \). It is the \( k \)th column of \( (L^\dagger)^\alpha \).

An important property of polyharmonic functions on \( \mathbb{R}^d \) is that they are conditionally positive definite, which means that one can use them to construct interpolants to scattered data. Here we state this result for our polyharmonic splines on graphs.

**Proposition 2.2.** Given a graph \( G \), a collection of distinct vertices \( V_1 = \{v_{kn}\}_{n=0}^{N_1-1} \) with \( N_1 \in \mathbb{Z}_{\geq 1} \), and data \( F = \{f_{km}\}_{m=0}^{N_1-1} \subset \mathbb{R} \), we can form the interpolant
\[ s_{F, V_1}(v) = C + \sum_{n=0}^{N_1-1} \beta_{kn} \Phi_\alpha(v, v_k), \]  
(5)

satisfying
\[ f_{km} = C + \sum_{n=0}^{N_1-1} \beta_{kn} \Phi_\alpha(v_{km}, v_k), \]  
(6)

for \( 0 \leq m \leq N_1 - 1 \). The vector of coefficients \( \beta_{kn} \) is orthogonal to constant vectors.

**Proof.** The Laplacian has a single zero eigenvalue with corresponding constant eigenvector. The same is true for the pseudo-inverse of the Laplacian. Also note that the pseudo-inverse is conditionally positive definite. Hence the submatrix corresponding to the rows and columns associated with \( V_1 \) is also conditionally positive definite.

Note that the coefficients \( \beta_{kn} \) and constant \( C \) are found by solving a matrix equation involving an augmented interpolation matrix \[19\] Section 8.5.
2.1 Variational characterization

Another important property of conditionally positive definite functions is a variational characterization. In particular, an interpolant of the form (5) should have the smallest norm over the collection of all interpolants in a particular semi-Hilbert space. The semi-Hilbert space in question is the reproducing kernel Hilbert space (or native space) associated with the kernel \((L^\dagger L)^\alpha\).

**Definition 2.3.** The native space \(N_\alpha\) for the kernel \((L^\dagger L)^\alpha\) is defined by the semi-inner product

\[
\langle f, g \rangle_{N_\alpha} = \langle L^{\alpha/2} f, L^{\alpha/2} g \rangle_{\ell_2(G)},
\]

where \(\langle \cdot, \cdot \rangle_{\ell_2(G)}\) is the standard inner product on \(\ell_2(G)\). The semi-norm for \(N_\alpha\) is denoted as \(|\cdot|_{H_\alpha^2(G)}\).

**Proposition 2.4.** Under the assumptions of proposition 2.2, the interpolant \(s_{F,\mathcal{V}_1}\) has minimal native space norm over all interpolants in \(N_\alpha\).

*Proof.* The proof of this result is analogous to that of [19, Theorem 13.2]. We provide the details for the benefit of the reader.

First, suppose \(g : \mathcal{V} \to \mathbb{R}\) is 0 on \(\mathcal{V}_1\). Since the vector \(\beta\) is orthogonal to constants,

\[
\langle g, s_{F,\mathcal{V}_1} \rangle_{N_\alpha} = \left\langle g, \sum_{n=0}^{N_1-1} \beta_{k_n} \Phi_\alpha(\cdot, v_{k_n}) \right\rangle_{N_\alpha}
= \sum_{n=0}^{N_1-1} \beta_{k_n} g(v_{k_n})
= 0.
\]

Hence for any interpolant \(s \in N_\alpha\),

\[
|s_{F,\mathcal{V}_1}|_{N_\alpha}^2 = \langle s_{F,\mathcal{V}_1}, s_{F,\mathcal{V}_1} \rangle_{N_\alpha}
= \langle s_{F,\mathcal{V}_1}, s_{F,\mathcal{V}_1} - s + s \rangle_{N_\alpha}
= \langle s_{F,\mathcal{V}_1}, s \rangle_{N_\alpha}
\leq |s_{F,\mathcal{V}_1}|_{N_\alpha} |s|_{N_\alpha}.
\]

3 Polyharmonic Lagrange functions

We are now in a position to define the polyharmonic Lagrange functions, which are the primary objects of study in this paper. Similar to the classical polynomial Lagrange functions, they are bases for constructing interpolants at a collection of vertices \(\mathcal{V}_1 := \{v_{k_n}\}_{n=1}^{N_1}\). The Lagrange function \(\chi(\cdot, v_{k_{n_0}})\) centered at \(v_{k_{n_0}}\) is a linear combination of polyharmonic splines

\[
\chi(\cdot, v_{k_{n_0}}) = C + \sum_{n=0}^{N_1-1} \beta_{k_n} \Phi_\alpha(\cdot, v_{k_n}).
\]

that satisfies

\[
\chi(v, v_{k_{n_0}}) = \begin{cases} 1, & v = v_{k_{n_0}} \\ 0, & v \in \mathcal{V}_1 \setminus \{v_{k_{n_0}}\}. \end{cases}
\]
### 3.1 Decay estimates for Lagrange functions

Our interest in Lagrange functions is based on their usefulness in constructing approximations to functions \( f : \mathcal{V} \rightarrow \mathbb{R} \) that are only known on a subset \( \mathcal{V}_1 \subset \mathcal{V} \). In particular,

i) The form of the approximation is very simple:

\[
\sum_{v \in \mathcal{V}_1} f(v) \chi(\cdot, v)
\]

(17)

ii) As we shall see, the polyharmonic Lagrange functions are well localized, as are their continuous counterparts [10].

In applications, localized bases are important for several reasons. For example, errors in the acquired sample data are confined to a small region. Also, if new data is acquired, the approximation can be updated locally.

In order to establish the decay of the Lagrange functions, we use a bulk chasing argument similar to the one used in [11]. The estimate, stated in theorem 3.7, will be based on the following additional assumptions on a graph \( \mathcal{G} \).

**Assumption 3.1.** We assume that the graph \( \mathcal{G} = \{\mathcal{V}, E \subset \mathcal{V} \times \mathcal{V}, w, \rho\} \) has \( N \in \mathbb{Z}_{\geq 2} \) vertices, i.e. \( \mathcal{V} = \{v_k\}_{n=0}^{N-1} \). The distance between adjacent vertices is at most \( \rho_{\text{max}} > 0 \). We also assume that the degree of the vertices is bounded by \( M \in \mathbb{Z}_{\geq 1} \).

**Remark 3.2.** Our goal is to provide a bound on the decay of the Lagrange functions for \((L^l)^2\) that is independent of \( N \). We therefore impose the uniform bounds \( \rho_{\text{max}} \) and \( M \) on the distances and degrees, respectively.

**Assumption 3.3.** Let \( \mathcal{G} \) satisfy the hypotheses of assumption 3.1. Let \( \mathcal{V}_1 \subset \mathcal{V} \) with fill distance \( h > 0 \); i.e.

\[
h := \max_{v \in \mathcal{V}} \min_{\tilde{v} \in \mathcal{V}_1} \rho(v, \tilde{v}),
\]

(18)

and let \( f : \mathcal{V} \rightarrow \mathbb{R} \) be zero on \( \mathcal{V}_1 \). If \( \mathcal{V}_2 \subset \mathcal{V} \) and

\[
\mathcal{V}_3 := \{v \in \mathcal{V} | \rho(v, \mathcal{V}_2) \leq 2h\},
\]

(19)

then there is a constant \( C > 0 \) such that

\[
\|f\|_{\ell_2(\mathcal{V}_2)}^2 \leq C |f|_{H^2_2(\mathcal{V}_3)}^2.
\]

(20)

Essentially, this assumption is saying that for functions with lots of zeros, the \( \ell_2 \) norm is controlled by the Sobolev semi-norm. This is known in the continuous domain (e.g. [12, Section 7.4]), and can be verified for graphs when the function \( f \) has a sufficient number of zeros. This assumption is valid for a sufficiently rich collection of graphs. Below, it is verified for certain classes of graphs. Moreover, a non-local version of this result for more general graphs is provided in appendix A. Let us also point out that without the assumption on the zeros of \( f \), eq. (20) is referred to as a Poincare inequality in [13]. There the author gives several estimates for such inequalities on various type of graphs.

**Proposition 3.4.** assumption 3.3 is valid for functions on cycle graphs.
Proof. Let $G$ be a cycle graph satisfying assumption 3.1 with $N \geq 3$. Assume the vertices are labeled such that $v_0$ is adjacent to $v_1$, $v_1$ is adjacent to $v_2$, ..., $v_{N-2}$ is adjacent to $v_{N-1}$ and $v_{N-1}$ is adjacent to $v_0$. Denote the vertices of $V_1$ as $\{v_{n_k}\}_{k=0}^{N_1-1}$ where $n_k < n_{k+1}$ for all $k$, and let $f$ be zero on $V_1$.

The idea for the proof is to cover $V_2$ (in fact we cover $G$ to remove the dependence on $V_2$) by subgraphs $G_k$, and show that the inequality holds on these smaller regions. We define the subgraphs so that they do not overlap too much, and therefore the inequality holds on unions of the $G_k$.

For simplicity, suppose $N_1$ is even; our argument can be suitably modified otherwise. Define the following induced subgraphs (which are paths connecting every other zero of $f$):

$$G_0 = \{v_{n_0}, v_{n_0+1}, \ldots, v_{n_2}\}$$

$$G_1 = \{v_{n_1}, v_{n_1+1}, \ldots, v_{n_3}\}$$

$$\vdots$$

$$G_{N_1-3} = \{v_{n_{N_1-3}}, v_{n_{N_1-3}+1}, \ldots, v_{n_{N_1-1}}\}$$

$$G_{N_1-2} = \{v_{n_{N_1-2}}, v_{n_{N_1-2}+1}, \ldots, v_{N-1}, v_0, \ldots, v_{n_0}\}$$

$$G_{N_1-1} = \{v_{n_{N_1-1}}, v_{n_{N_1-1}+1}, \ldots, v_{N-1}, v_0, \ldots, v_{n_1}\}.$$

The length of each path $G_k$ is at most $4h$. For any given set $V_2$, there is a union of $G_k$'s such that the union of their interiors (denoted $\text{int}(G_k)$) cover $V_2$ and is contained in $V_3$. Let $K$ be the index set for such a collection; i.e.,

$$V_2 \subseteq \bigcup_{k \in K} \text{int}(G_k) \subseteq V_3$$

Also, note that each vertex is in at most two paths $\text{int}(G_k)$.

The subgraphs $G_k$ were constructed so that the function $f$ satisfies Dirichlet boundary conditions on each one. For each $k$, consider the submatrix $L_{\text{int}(G_k)}$ of the Laplacian, where the rows and columns correspond to the vertices of $\text{int}(G_k)$. The minimal eigenvalue of this matrix, is referred to as a Dirichlet eigenvalue [2], and note that it is positive. We denote the Dirichlet eigenvalue on $\text{int}(G_k)$ as $\lambda_{0}^{G_k}$.

We now have

$$\|f\|_{\ell_2(V_2)}^2 \leq \sum_{k \in K} \|f\|_{\ell_2(\text{int}(G_k))}^2$$

$$\leq \sum_{k \in K} \left(\frac{1}{\lambda_{0}^{G_k}}\right)^2 \|Lf\|_{\ell_2(\text{int}(G_k))}^2$$

$$\leq \min_{0 \leq k \leq N_1-1} \left\{\left(\frac{1}{\lambda_{0}^{G_k}}\right)^2\right\} \sum_{k \in K} \|Lf\|_{\ell_2(\text{int}(G_k))}^2$$

$$\leq 2 \min_{0 \leq k \leq N_1-1} \left\{\left(\frac{1}{\lambda_{0}^{G_k}}\right)^2\right\} \|Lf\|_{\ell_2(V_3)}^2.$$

In the proof of the previous proposition, we derived a constant $C$ for (20). The constant depends only on the Dirichlet eigenvalues for the subgraphs $G_k$, which were constructed to fit the
density of the vertices $V_1$ in $V$. In particular, this means that the constant does not depend on the size of the graph. For example, if a cycle graph has equally weighted edges and bounded fill distance $h$, then the constant $C$ will have a bound that is independent of the number of vertices $N$. In this case, the Dirichlet eigenvalues are actually known, cf. Lemma I.14 of [5].

In order to verify assumption 3.3 we believe that it is necessary to have a covering of the graph $G$ by subgraphs satisfying some boundary conditions. In proposition 3.4 the subgraphs satisfy Dirichlet boundary condition. This is particularly relevant to learning applications where training data sets are larger than the collection of unknowns being evaluated. Also note that one could consider other boundary conditions such as Neumann boundary conditions and apply lemma A.1 to each subgraph. The assumption would also be true if a subgraph satisfied a combination of Dirichlet and Neumann boundary conditions. The next proposition describes how assumption 3.3 is satisfied by graphs constructed for machine learning.

**Proposition 3.5.** Let $G$ be a graph as in assumption 3.1. Suppose the vertices of $G$ are divided into two sets $V_k$ and $V_u$ (corresponding to known and unknown values for some function on $G$), and suppose that there are no edges connecting distinct vertices in $V_u$. If the length of the edges in $G$ are bounded below by $\rho_{\text{max}}/2$ and $\rho = w^{-1}$, then $G$ satisfies assumption 3.3.

**Proof.** The idea for the proof is the same as for cycle graphs. We need to define a cover of $G$ by subgraphs that have bounded Dirichlet eigenvalues. Moreover, the subgraphs cannot be too large; in addition to covering $V_2$, they must lie within $V_3$. This is the reason for the restriction on the edge distances. The fact that the minimum eigenvalue of the Laplacian on these subgraphs is bounded will follow from a bound on the Cheeger constant [7, Chapter 3].

We construct a subgraph, starting from a given vertex $v_0$ and grow the subgraph until it is surrounded by vertices from $V_k$. We introduce the notation $\mathcal{N}_v$ for the neighbors of a vertex $v$. To make our argument precise, let $v_0 \in V$, and consider the subgraph $G_{v_0}$ consisting of the vertices

$$\Omega_{v_0} := \{v_0\} \cup \mathcal{N}_{v_0} \bigcup_{v \in \mathcal{N}_{v_0} \cap V_u} \mathcal{N}_v,$$

i.e. $\Omega_{v_0}$ consists of the neighbors of $v_0$ and the neighbors of the neighbors that are in $V_u$. As no two vertices of $V_u$ are connected by an edge, the boundary of int($G_{v_0}$) consists entirely of vertices from $V_k$.

To bound the Cheeger constant, and hence the minimum eigenvalue, on int($G_{v_0}$), we note the following:

a) The number of vertices in $G_{v_0}$, depending on the maximum degree $M$, is at most $1 + M^2$

b) Given a nonempty subset $U$ of int($G_{v_0}$), the sum of the edge weights between $U$ and its boundary in $G$ is at least $\rho_{\text{max}}^{-1}$. Also, the sum of the edge weights (counting multiplicities) over the vertices of $U$ is at most $(M + 1)M2\rho_{\text{max}}^{-1}$.

Using these facts, we find a lower bound for the Cheeger constant for subgraphs of this type [7, Chapter 3]. In particular the Cheeger constant $\tilde{h}$ satisfies

$$\tilde{h}(\text{int}(G_v)) \geq \frac{\rho_{\text{max}}^{-1}}{(M + 1)M2\rho_{\text{max}}^{-1}} \geq \frac{1}{(M + 1)M2}$$

(33)
Cheeger’s inequality then implies that the minimum eigenvalue $\lambda_{0}^{G_{v}}$ of $L_{\text{int}(G_{v})}$ satisfies
\begin{equation}
\lambda_{0}^{G_{v}} \geq \frac{1}{2} h(\text{int}(G_{v}))^{2} \geq \frac{1}{8} \left(\frac{1}{(M+1)M}\right)^{2},
\end{equation}
so
\begin{equation}
\left(\frac{1}{\lambda_{0}^{G_{v}}}\right)^{2} \leq 64(M+1)^{4}M^{4}.
\end{equation}

Now, given a set $V_{2}$, we cover it by
\[ \bigcup_{v \in V_{2}} \text{int}(G_{v}) \subseteq V_{3}. \]
Containment in $V_{3}$ follows from the fact that $\rho_{\text{max}} \leq 2(\rho_{\text{max}}/2) \leq 2h$. We also note that the construction of $G_{v_{0}}$ means that any given vertex $v \in G$ can be in the interior of at most $M$ subgraphs of this type, and the number of vertices in $G_{v_{0}}$, depending on the maximum degree $M$, is at most $1 + M^{2}$.

We now finish the proof. Given any function $f$ that is zero on $V_{k}$, we have
\begin{equation}
\|f\|_{\ell_{2}(V_{2})}^{2} \leq \sum_{v \in V_{2}} \|f\|_{\ell_{2}(\text{int}(G_{v}))}^{2} \leq \sum_{v \in V_{2}} \left(\frac{1}{\lambda_{0}^{G_{v}}}\right)^{2} \|Lf\|_{\ell_{2}(\text{int}(G_{v}))}^{2} \leq 64(M+1)^{4}M^{4} \sum_{v \in V_{2}} \|Lf\|_{\ell_{2}(\text{int}(G_{v}))}^{2} \leq 64(M+1)^{4}M^{5} \|Lf\|_{\ell_{2}(V_{3})}^{2}.
\end{equation}

We now proceed to a key lemma for deriving the decay of the Lagrange functions.

**Lemma 3.6.** Let $G = \{\mathcal{V}, E \subset \mathcal{V} \times \mathcal{V}, w, \rho\}$ satisfy the hypotheses of assumption 3.1 and assumption 3.3. Let $v_{n_{0}} \in V_{1} \subset \mathcal{V}$, and consider the Lagrange function $\chi(\cdot, v_{n_{0}})$ associated with $(L^{1})^{2}$ and the vertex set $V_{1}$ with fill distance $h$. If $0 < 3\rho_{\text{max}} + 2h < r_{2} < r_{3} < \infty$, then there is a constant $\mu < 1$ such that
\begin{equation}
|\chi|_{H_{2}^{2}(G \setminus B(v_{n_{0}}, r_{4}))} \leq \mu |\chi|_{H_{2}^{2}(G \setminus B(v_{n_{0}}, r_{1}))},
\end{equation}
where $r_{1} := r_{2} - 2\rho_{\text{max}} - 2h$ and $r_{4} := r_{3} + 2h$.

**Proof.** Our proof is similar to that of [9, 8]. In this proof, all balls and annuli are centered at $v_{n_{0}}$. We simplify our notation by omitting the center: for example we write $B(r_{0})$ rather than $B(v_{n_{0}}, r_{0})$.

Let $\phi$ be a function satisfying
\begin{itemize}
    \item $0 \leq \phi(v) \leq 1$
\end{itemize}
• \( \phi(v) = 1 \) for \( v \in B(r_2) \)

• \( \phi(v) = 0 \) for \( v \notin B(r_3) \)

Recall the interpolation criterion for \( \chi \)

\[
\chi(v, v_n) = \begin{cases} 
1, & v = v_n \\
0, & v \in \mathcal{V}_1 \setminus \{v_n\}
\end{cases},
\]

and notice that the product \( \phi \chi \) satisfies the same interpolation conditions. Hence the variational property of \( \chi \) implies

\[
|\chi|^2_{H^2(G)} \leq |\phi \chi|^2_{H^2(G)} = |\chi|^2_{H^2(B(r_2 - \rho_{max}))} + |\phi \chi|^2_{H^2(A(r_2 - \rho_{max}, r_3))}.
\]

Using the estimate above, we have

\[
|\chi|^2_{H^2(G\setminus B(r_2))} = |\chi|^2_{H^2(G)} - |\chi|^2_{H^2(B(r_2))} \leq \left(|\chi|^2_{H^2(B(r_2 - \rho_{max}))} + |\phi \chi|^2_{H^2(A(r_2 - \rho_{max}, r_3))}\right) - |\chi|^2_{H^2(B(r_2))}.
\]

Simplifying the expression gives

\[
|\chi|^2_{H^2(G\setminus B(r_2))} \leq \left(|\chi|^2_{H^2(B(r_2 - \rho_{max}))} - |\chi|^2_{H^2(B(r_2))}\right) + |\phi \chi|^2_{H^2(A(r_2 - \rho_{max}, r_3))} \leq |\phi \chi|^2_{H^2(A(r_2 - \rho_{max}, r_3))}.
\]

Using the definition of the semi-norm and properties of \( L \),

\[
|\chi|^2_{H^2(G\setminus B(r_2))} \leq \|L(\phi \chi)\|^2_{L^2(A(r_2 - \rho_{max}, r_3))} \leq 4 \|\phi \chi\|^2_{L^2(A(r_2 - 2\rho_{max}, r_3))} \leq 4 \|\chi\|^2_{L^2(A(r_2 - 2\rho_{max}, r_3))} \leq 4 \|\chi\|^2_{L^2(A(r_2 - 2\rho_{max}, r_3))}.
\]

Now we apply assumption \( 3.3 \) with \( f = \chi \), \( \mathcal{V}_2 = A(r_2 - 2\rho_{max}, r_3) \), and \( \mathcal{V}_3 = A(r_2 - 2\rho_{max} - 2h, r_3 + 2h) = A(r_1, r_4) \). This implies that there is a constant \( C_0 \) such that

\[
|\chi|^2_{H^2(G\setminus B(r_1))} \leq C_0 |\chi|^2_{H^2(A(r_1, r_4))}.
\]

Writing the annulus as a set difference

\[
|\chi|^2_{H^2(G\setminus B(r_1))} \leq C_0 \left(|\chi|^2_{H^2(G\setminus B(r_1))} - |\chi|^2_{H^2(G\setminus B(r_4))}\right),
\]

which implies

\[
(1 + C_0) |\chi|^2_{H^2(G\setminus B(r_4))} \leq C_0 |\chi|^2_{H^2(G\setminus B(r_1))}.
\]

This is equivalent to

\[
|\chi|^2_{H^2(G\setminus B(r_4))} \leq \frac{C_0}{1 + C_0} |\chi|^2_{H^2(G\setminus B(r_1))}.
\]
Theorem 3.7. Under the assumptions of lemma 3.6, the Lagrange function $\chi(\cdot, v_{n_0})$ satisfies exponential decay away from its center. In particular, there are constants $C, T > 0$ such that

$$|\chi(v_{n_1}, v_{n_0})| \leq C \mu^T \rho(v_{n_1}, v_{n_0}),$$

(58)

where $\mu < 1$ comes from lemma 3.6.

Proof. Let $v_{n_0} \in V_1$ and $v_{n_1} \in V$ such that $\rho(v_{n_0}, v_{n_1})$ satisfies

$$M(4h + 3\rho_{\max}) + 2h + 2\rho_{\max} \leq \rho(v_{n_0}, v_{n_1}) \leq M(4h + 3\rho_{\max}) + 2h + 3\rho_{\max}$$

(59)

for some positive integer $M$. We now define a sequence of balls, centered at $v_{n_0}$, whose radii are shrinking by an amount prescribed by lemma 3.6

$$B_k := B(v_{n_0}, (M - k)(4h + 3\rho_{\max}) + 2h + 2\rho_{\max})$$

(60)

for $k = 0, \ldots, M$.

Then

$$|\chi(v_{n_1}, v_{n_0})| \leq C |\chi(v_{n_0})|_{H^2_2(B(v_{n_1}, 2h))}$$

(61)

$$\leq C |\chi(v_{n_0})|_{H^2_2(G \setminus B_1)}$$

(62)

Applying lemma 3.6

$$|\chi(v_{n_1}, v_{n_0})| \leq C \mu |\chi(v_{n_0})|_{H^2_2(G \setminus B_1)}$$

(63)

$$\leq C \mu^M |\chi(v_{n_0})|_{H^2_2(G \setminus B_M)}$$

(64)

$$\leq C \mu^{\rho(v_{n_1}, v_{n_0}) - 2h - 3\rho_{\max}} |\chi(v_{n_0})|_{H^2_2(G)}$$

(65)

Let us point out that, in addition to exponential decay, we have shown more precisely how the Lagrange functions decay in terms of the fill distance $h$ and the maximum distance between points $\rho_{\max}$. This can be seen in (65). In machine learning applications, $h$ and $\rho_{\max}$ will generally decrease as more data is obtained and added for training, leading to more localized basis functions that. This is analogous to B-splines on euclidean spaces that scale with the density of the sampling grid.

On infinite graphs, the Moore-Penrose pseudo-inverse of the Laplacian is known to satisfy exponential decay [18, Theorem 5.1], and exponential decay of our Lagrange functions could be derived accordingly in that setting. However, we would not be able to obtain the more precise decay of theorem 3.7. In fact, it is clear from the examples in the next section that the Lagrange functions on graphs are much more localized than the polyharmonic functions, just as in the case of polyharmonic functions on continuous domains.

As a corollary of our theorem, we can deduce the exponential decay of the coefficients of the Lagrange functions. This follows from the next proposition that relates the coefficients to a native space semi-innerproduct between two Lagrange functions.
Proposition 3.8. Let \( v_{n_0}, v_{n_1} \in \mathcal{V}_1 \subset \mathcal{V} \) and consider the Lagrange functions associated with \( L^\alpha \) for \( \alpha > 0 \)

\[
\chi(\cdot, v_{n_0}) = C_0 + \sum_{v \in \mathcal{V}_1} \gamma_v \Phi_\alpha(\cdot, v)
\]

(66)

\[
\chi(\cdot, v_{n_1}) = C_1 + \sum_{v \in \mathcal{V}_1} \beta_v \Phi_\alpha(\cdot, v)
\]

(67)

Then we have

\[
\langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\mathcal{N}_\alpha} = \gamma_{v_{n_1}} = \beta_{v_{n_0}}
\]

(68)

Proof. This follows from the definition of the native space norm and the symmetry of the Laplacian.

\[
\langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\mathcal{N}_\alpha} = \langle L^\alpha \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\ell_2(\mathcal{G})}
\]

(69)

\[
= \left\langle L^\alpha \left( C_0 + \sum_{v \in \mathcal{V}_1} \gamma_v \Phi_\alpha(\cdot, v) \right), \chi(\cdot, v_{n_1}) \right\rangle_{\ell_2(\mathcal{G})}
\]

(70)

\[
= \left\langle \sum_{v \in \mathcal{V}_1} \gamma_v L^\alpha \Phi_\alpha(\cdot, v), \chi(\cdot, v_{n_1}) \right\rangle_{\ell_2(\mathcal{G})}
\]

(71)

\[
= \left\langle \sum_{v \in \mathcal{V}_1} \gamma_v e_v, \chi(\cdot, v_{n_1}) \right\rangle_{\ell_2(\mathcal{G})}
\]

(72)

\[
= \sum_{v \in \mathcal{V}_1} \gamma_v \langle e_v, \chi(\cdot, v_{n_1}) \rangle_{\ell_2(\mathcal{G})}
\]

(73)

\[
= \gamma_{v_{n_1}}
\]

(74)

By symmetry, this is also equal to \( \beta_{v_{n_0}} \).

\[\square\]

Corollary 3.9. The coefficients of the Lagrange functions decay exponentially

Proof. Let \( v_{n_0}, v_{n_1} \in \mathcal{V}_1 \). Considering proposition 3.8 to show exponential decay of the coefficients, it suffices to show the exponential decay of the semi-inner product \( \langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\mathcal{N}_2(\mathcal{G})} \).

Now, suppose

\[
2M(4h + 3\rho_{\max}) + 2h + 2\rho_{\max} \leq \rho(v_{n_0}, v_{n_1}) \leq 2M(4h + 3\rho_{\max}) + 2h + 3\rho_{\max}
\]

(75)

Consider decomposing the graph into the following sets based on the distance

\[
r = M(4h + 3\rho_{\max}) + 2h + 2\rho_{\max}.
\]

We define the sets

\[
X_0 = \{ v \in \mathcal{V} \mid \rho(v, v_{n_0}) \leq r \}
\]

(76)

\[
X_1 = \{ v \in \mathcal{V} \mid \rho(v, v_{n_1}) > r \}
\]

(77)

Then

\[
\langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\mathcal{N}_2(\mathcal{G})} = \langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\mathcal{N}_2(X_0)} + \langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{\mathcal{N}_2(X_1)}
\]

(78)

\[
\leq |\chi(\cdot, v_{n_0})|_{H^2_\alpha(X_0)} |\chi(\cdot, v_{n_1})|_{H^2_\alpha(X_0)} + |\chi(\cdot, v_{n_0})|_{H^2_\alpha(X_1)} |\chi(\cdot, v_{n_1})|_{H^2_\alpha(X_1)}.
\]

(79)
Let $\mu_0, \mu_1 < 1$ be the constants corresponding to the decay of $\chi(\cdot, v_{n_0})$ and $\chi(\cdot, v_{n_1})$ respectively. Also, define $\mu = \max\{\mu_0, \mu_1\}$. Then

\[
\langle \chi(\cdot, v_{n_0}), \chi(\cdot, v_{n_1}) \rangle_{N_2(G)} \leq |\chi(\cdot, v_{n_0})|_{H^2_2(G_0)} \mu_1^M + \mu_0^M |\chi(\cdot, v_{n_1})|_{H^2_2(G_1)} \tag{80}
\]

\[
\leq C \mu \frac{\rho(v_{n_0}, v_{n_1})^{2h - 3\rho_{\max}}}{4h + 3\rho_{\max}} \tag{81}
\]

\[
4 \text{ Local Lagrange functions}
\]

Given the exponential decay of the Lagrange coefficients, we can truncate the expansion to form truncated Lagrange functions that closely approximate the full Lagrange functions. While these functions have a simpler form than the Lagrange functions, they require the same amount of computation to construct. The purpose of the truncated Lagrange functions is to provide a theoretical link between the Lagrange and local Lagrange functions defined below. Ultimately, we intend to use the more computationally efficient local Lagrange functions.

The Lagrange function $\chi(\cdot, v_{n_0})$ centered at $v_{n_0}$ is a linear combination of polyharmonic splines

\[
\chi(\cdot, v_{k_{n_0}}) = C + \sum_{v \in V} \beta_v \Phi_\alpha(\cdot, v). \tag{82}
\]

We form the truncated Lagrange function $\tilde{\chi}(\cdot, v_{k_{n_0}})$ in terms of a distance $K > 0$.

In particular, we define a neighborhood $B(v_{n_0}, K)$ and consider

\[
\tilde{\chi}(\cdot, v_{k_{n_0}}) = C + \sum_{v \in B(v_{n_0}, K)} \left( \beta_v - \frac{1}{|B(v_{n_0}, K)|} \sum_{w \in B(v_{n_0}, K)} \beta_w \right) \Phi_\alpha(\cdot, v). \tag{83}
\]

The modification on the coefficients ensures that they are orthogonal to constant vectors (the null space of the local interpolation matrix).

The local Lagrange functions $\bar{\chi}(\cdot, v_{k_{n_0}})$ are constructed analogously to the Lagrange functions; however, they only use basis functions centered at points of $B(v_{n_0}, K)$ and the interpolation conditions are only enforced on this ball.

**Definition 4.1.** Given a neighborhood $B(v_{n_0}, K)$ of $v_{n_0} \in V$ with $K > 0$, we define the local Lagrange function

\[
\bar{\chi}(\cdot, v_{k_{n_0}}) = \tilde{C} + \sum_{v \in B(v_{n_0}, K)} \tilde{\beta}_v \Phi_\alpha(\cdot, v), \tag{84}
\]

where

\[
\bar{\chi}(v, v_{k_{n_0}}) = \begin{cases} 1, & v = v_{k_{n_0}} \\ 0, & v \in B(v_{n_0}, K) \setminus \{v_{k_{n_0}}\} \end{cases} \tag{85}
\]

and the vector of coefficients $\tilde{\beta}_v$ is orthogonal to constant vectors.

**Conjecture 4.2.** The local Lagrange functions can be made arbitrarily close to the full Lagrange functions with a relatively small $K$ value.

This conjecture has been established for basis functions on continuous domains [6], and we experimentally establish the connection in the next section.
5  Examples and simulations

Here we illustrate the localization theorems and apply the local Lagrange functions in classification problems. First, we compare the Green’s function with the Lagrange function in fig. 1 for a cycle graph with 256 nodes, where every fourth vertex is an interpolation node. In fig. 2 we compare the basis functions on a lattice graph. In both cases, we see that the Lagrange functions exhibit much faster decay and are much better localized than the corresponding Green’s functions.

5.1  Experiment: Wine quality

One potential application of our proposed basis functions is kernel-based machine learning on graphs. The graphs considered are ideal candidates to satisfy assumption 3.3 as a majority of the data sites are known values. We illustrate this with the following example of classifying wine quality based on eleven numeric parameters: sulphates, alcohol, residual sugar, citric acid, total sulfur dioxide, free sulfur dioxide, volatile acidity, density, pH, chlorides, and fixed acidity. The data set comes from the UC Irvine Machine Learning Repository [4].

The algorithm for classification is the following. We construct a graph of all the instances
(known and unknown); each instance is a vertex. Edges are defined between vertices whose parameters are close. We bound the maximum number of edges coming out of a vertex. The edges are weighted based on the distances between the parameters. Then we compute the Lagrange functions at the known data sites. Using these basis functions, we interpolate to compute the class for unknown data sites.

The algorithm is quite straightforward and currently does not incorporate methods to identify important parameters (or combinations of parameters). However, even with this naive approach, our algorithm is still competitive with the more common models. We believe that this is a good indication of the potential of the approach, and we are currently working on a refined algorithm that can handle more complex data sets.

For the simulation, we split the data set into five groups for a 5-fold validation. For a single test, one of the groups was considered unknown while the other four groups were used as known data. After five tests, the results are combined to determine the precision of the algorithm. We ran the simulation twenty times, randomizing the data each time. The displayed results are an average over these twenty simulations. This experiment is analogous to that of [3], which we use for comparison. We recall only the best method (SVM) from that paper, which compares three different algorithms. One distinction to note is that we do not optimize the parameters; we only normalize each parameter to have mean zero and standard deviation one. The results of this experiment are displayed in tables [1] and [2]. Our method shows improvement over the other results as we have an average precision of 67.2% for the red wine data set and 67.4% on the white wine data set. In comparison, the precision of the SVM method is 62.4% and 64.6%, respectively.

Table 1: Red wine data set. Over 20 simulations with 5-fold validation and a tolerance of 0.5, we have the following results. The confusion matrix and precision for each class are displayed. The SVM results are from [3].

| Actual Class | 4 | 5 | 6 | 7 | 8 |
|--------------|---|---|---|---|---|
| 3            | 1 | 7 | 1 | 0 | 0 |
| 4            | 0 | 31| 20| 0 | 0 |
| 5            | 3 | 515|156| 5 | 0 |
| 6            | 0 | 145|456| 36| 0 |
| 7            | 0 | 7 | 91|99 | 0 |
| 8            | 0 | 0 | 11| 4 | 1 |

Precision: RBF 13.5 72.9 61.9 68.3 100.0  
SVM 20.0 67.5 57.7 58.6 0.0

6 Discussion

We have introduced the analog of radial basis functions on graphs and verified properties that are known to hold on continuous domains. Our main result is the decay of the Lagrange functions. Our approach was a spatial domain estimate, rather than the Fourier-based results that are commonly used for uniform data. This bulk-chasing argument was adapted from the one used on continuous domains for non-uniform data. This connection is encouraging and leads us to
Table 2: White wine data set. Over 20 simulations with 5-fold validation and a tolerance of 0.5, we have the following results. The confusion matrix and precision for each class are displayed. The SVM results are from [3].

| Actual Class | White wine predictions |
|--------------|------------------------|
|              | 4 | 5 | 6 | 7 | 8 |
| 3            | 1 | 9 | 5 | 3 | 0 |
| 4            | 29 | 82 | 46 | 5 | 0 |
| 5            | 7 | 981 | 443 | 24 | 0 |
| 6            | 2 | 306 | 1700 | 186 | 1 |
| 7            | 0 | 15 | 334 | 523 | 5 |
| 8            | 0 | 2 | 46 | 57 | 68 |
| 9            | 0 | 0 | 1 | 2 | 0 |

Precision: RBF 71.2 70.1 65.9 65.2 90.9
SVM 63.3 72.6 60.3 67.8 85.5

believe that additional properties will also carry over. A future goal is to show that quasi-interpolating local-Lagrange functions satisfy analogous localization [6]. This is important as it reduces the computational cost of computing the basis functions. The construction of such bases raises questions about how to properly truncate coefficients and the basis functions themselves. Other important properties such as stability with respect to $\ell_p$ are also of interest. Establishing these results should benefit the continuous domain theory as the graph setting is more precise in terms of bounding constants such as $\mu$ of theorem 3.7 as well as constants appearing in the footprint radius of local Lagrange functions.

The potential applications of such bases include kernel-based machine learning algorithms, where data is well represented using a graph framework. Preliminary results are promising and show comparable results to commonly used methods. Also, we expect improved results by adding a refined training to our algorithm.

A Sobolev functions with zeros

Lemma A.1. Let $L$ denote the Laplacian on a connected $G = \{V, E, w\}$, with $N$ vertices. Let its eigenvectors and eigenvalues be denoted as $\{\Lambda_n\}_{n=0}^{N-1}$, $\{\lambda_n\}_{n=0}^{N-1}$. Let $f : V \rightarrow \mathbb{R}$ be a function satisfying $f(v_{n_0}) = 0$ for some vertex $v_{n_0} \in V$. Then

$$\|f\|_{\ell_2(G)} \leq \frac{\sqrt{N+1}}{(\lambda_1)^{\alpha/2}} |f|_{H^\alpha_2(G)}$$

Proof. Our proof is based on the proof of an analogous result in [12 Chapter 7]. First, we use Parseval’s identity to express the left-hand side as a sum of Fourier coefficients

$$\|f\|_{\ell_2}^2 = \sum_{n=0}^{N-1} |\widehat{f}(\lambda_n)|^2$$

$$= |\widehat{f}(\lambda_0)|^2 + \sum_{n=1}^{N-1} |\widehat{f}(\lambda_n)|^2.$$
Since \( f \) has a zero at \( v_{n_0} \)

\[
0 = f(v_{n_0}) = \hat{f}(\lambda_0)\Lambda_0(v_{n_0}) + \sum_{n=1}^{N-1} \hat{f}(\lambda_n)\Lambda_n(v_{n_0}).
\] (89)

We assume the eigenvectors are assume to have norm 1. In particular, \( \Lambda_0 \) is the constant vector \( \frac{1}{\sqrt{N}} \). Substituting this into the previous equation, we have

\[
\|f\|_{\ell_2}^2 = \sqrt{N} \left( \sum_{n=1}^{N-1} \left| \Lambda_n(v_{n_0}) \right|^2 \right) + \left( \sum_{n=1}^{N-1} \left| \hat{f}(\lambda_n) \right|^2 \right) + \sum_{n=1}^{N-1} \left| \hat{f}(\lambda_n) \right|^2.
\] (90)

Next, we apply the Cauchy-Schwartz inequality

\[
\|f\|_{\ell_2}^2 \leq N \left( \sum_{n=1}^{N-1} |\Lambda_n(v_{n_0})|^2 \right) \left( \sum_{n=1}^{N-1} |\hat{f}(\lambda_n)|^2 \right) + \sum_{n=1}^{N-1} \left| \hat{f}(\lambda_n) \right|^2.
\] (92)

Finally we simplify

\[
\|f\|_{\ell_2}^2 \leq N \left( \sum_{n=1}^{N-1} \left| \hat{f}(\lambda_n) \right|^2 \right) + \sum_{n=1}^{N-1} \left| \hat{f}(\lambda_n) \right|^2
\] (93)

\[
\leq (N + 1) \sum_{n=1}^{N-1} \left| \hat{f}(\lambda_n) \right|^2
\] (94)

\[
\leq \frac{N + 1}{(\lambda_1)^{\alpha}} \sum_{n=1}^{N-1} (\lambda_n)^{\alpha} \left| \hat{f}(\lambda_n) \right|^2
\] (95)

\[
= \frac{N + 1}{(\lambda_1)^{\alpha}} \|f\|_{H_2}^2.
\] (96)

References

[1] J.J. Benedetto and P.J. Koprowski. Graph theoretic uncertainty principles. In Sampling Theory and Applications (SampTA), 2015 International Conference on, pages 357–361. IEEE, 2015.

[2] F. R. K. Chung. Spectral graph theory, volume 92 of CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997.

[3] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4):547–553, 2009.

[4] D. Dheeru and E. Karra Taniskidou. UCI machine learning repository, 2017.
[5] R. B. Ellis III. Chip-firing games with dirichlet eigenvalues and discrete green’s functions, 2002.

[6] E. Fuselier, T. Hangelbroek, F. J. Narcowich, J. D. Ward, and G. B. Wright. Localized bases for kernel spaces on the unit sphere. *SIAM J. Numer. Anal.*, 51(5):2538–2562, 2013.

[7] Alexander Grigor’yan. *Introduction to Analysis on Graphs*, volume 71. American Mathematical Society, 2018.

[8] T. Hangelbroek. The penalized Lebesgue constant for surface spline interpolation. *Proc. Amer. Math. Soc.*, 140(1):173–187, 2012.

[9] T. Hangelbroek, F. J. Narcowich, and J. D. Ward. Kernel approximation on manifolds I: bounding the Lebesgue constant. *SIAM J. Math. Anal.*, 42(4):1732–1760, 2010.

[10] T. Hangelbroek, F. J. Narcowich, and J. D. Ward. Polyharmonic and related kernels on manifolds: interpolation and approximation. *Found. Comput. Math.*, 12(5):625–670, 2012.

[11] O. V. Matveev. Spline interpolation of functions of several variables and bases in Sobolev spaces. *Trudy Mat. Inst. Steklov.*, 198:125–152, 1992.

[12] F. Natterer. *The mathematics of computerized tomography*. B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester, 1986.

[13] I. Pesenson. Sampling in Paley-Wiener spaces on combinatorial graphs. *Trans. Amer. Math. Soc.*, 360(10):5603–5627, 2008.

[14] I. Pesenson. Variational splines and paley–wiener spaces on combinatorial graphs. *Constructive Approximation*, 29(1):1–21, Feb 2009.

[15] I. Pesenson. Removable sets and approximation of eigenvalues and eigenfunctions on combinatorial graphs. *Applied and Computational Harmonic Analysis*, 29(2):123 – 133, 2010.

[16] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[17] A.J. Smola and R. Kondor. Kernels and regularization on graphs. In *Learning theory and kernel machines*, pages 144–158. Springer, 2003.

[18] Q. Sun. Wiener’s lemma for infinite matrices. *Trans. Amer. Math. Soc.*, 359(7):3099–3123, 2007.

[19] H. Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.