

CONVOLUTIONAL FILTERING ON SAMPLED MANIFOLDS

Zhiyang Wang[†] Luana Ruiz^{*} Alejandro Ribeiro[†]

[†]Department of Electrical and Systems Engineering, University of Pennsylvania, PA
^{*} Simons-Berkeley Institute, CA

ABSTRACT

The increasing availability of geometric data has motivated the need for information processing over non-Euclidean domains modeled as manifolds. The building block for information processing architectures with desirable theoretical properties such as invariance and stability is convolutional filtering. Manifold convolutional filters are defined from the manifold diffusion sequence, constructed by successive applications of the Laplace-Beltrami operator to manifold signals. However, the continuous manifold model can only be accessed by sampling discrete points and building an approximate graph model from the sampled manifold. Effective linear information processing on the manifold requires quantifying the error incurred when approximating manifold convolutions with graph convolutions. In this paper, we derive a non-asymptotic error bound for this approximation, showing that convolutional filtering on the sampled manifold converges to continuous manifold filtering. Our findings are further demonstrated empirically on a problem of navigation control.

Index Terms— Manifold filter, manifold signal processing, Laplace-Beltrami operator, graph Laplacian

1. INTRODUCTION

Applications involving structured data in non-Euclidean domains are attracting more and more attention these days, including but not limited to robot coordination [1, 2], resource allocation over wireless networks [3] and 3D shape analysis [4, 5]. Graph convolutional filtering [6, 7] and manifold convolutional filtering [8, 9] have become the top choices to process information over these non-Euclidean domains. This can be attributed to the fact that the convolution operation consists of local diffusions over the geometric structure, which capture features that are shared across the whole domain. Moreover, convolutional filtering provides the fundamental block for constructing deep learning architectures, which are powerful models [10] to solve a wide range of geometric problems.

Unlike graphs, which are discrete objects, manifolds are not directly accessible in general because they are infinite-dimensional continuous spaces [4, 8]. Instead, we typically have access to a set of points sampled from the manifold, or a *sampled manifold*. In many manifold problems, a *graph* capturing both the local and global connection structure is built from this sampled manifold to approximate the underlying continuous manifold [3, 11] (herein, we use the terms *sampled manifold* and *graph* interchangeably). Convolutional filters constructed on graphs have been used to approximately process information over the manifold, which has been supported both experimentally and theoretically [8, 10, 12]. However, an explicit non-asymptotic theoretical result closing the gap between graph convolutional filtering and manifold convolutional filtering is still lacking.

Contributions. In this paper, we define manifold filtering as a convolutional operation based on diffusions that are defined as exponentials of the Laplace-Beltrami (LB) operator \mathcal{L} of the embedded manifold $\mathcal{M} \subset \mathbb{R}^N$ [8]. Graphs are constructed by sampling a set of points uniformly over the manifold, with edge weights given by application of a Gaussian kernel. Denoting the graph Laplacian \mathbf{L}_n^ϵ , we derive upper bounds for both the operator norm difference and the difference between the spectra of the graph Laplacian \mathbf{L}_n^ϵ and the LB operator \mathcal{L} . Then, to deal with the infinite dimensional spectrum and the sharp variations in the high frequency domain of LB operator \mathcal{L} , we import the definition of the Frequency Difference Threshold (FDT) filters (Definition 2) [13] to separate the spectrum into finite eigenvalue partitions. Finally, we show that when applying FDT filters to graphs sampled from the manifold, graph filtering converges to the manifold filtering with a non-asymptotic approximation bound in the order of $n^{-1/(2d+8)}$ (Theorem 3).

Related Works. Graph convolutional filtering has been thoroughly discussed in previous works [6, 7, 14]. A large number of works have studied the convergence of graph filters, such as [15] and [16] which consider graphs that are dense or relatively sparse, and which are both sampled from a graphon model. Unlike the graphon, the manifold is more realistic and can be seen as a limit for graphs that have finite degree [17]. Convolutional filtering in manifolds has been discussed in [8, 9, 12], with different definitions depending on how the graphs are sampled from manifolds. The connection between graph filtering and manifold filtering has been discussed in [8]. In this paper, it is proved that manifold filtering can recover graph filtering by discretization in both the space and time domains; and that filtering on graphs sampled from the manifold converges to manifold filtering. However, there are no theoretical convergence rates. In the present paper, we extend upon the results of [8] by proving a non-asymptotic approximation bound for graph filters sampled from manifold filters.

2. PRELIMINARY DEFINITIONS

In order to relate graph filters and manifold filters, we start by reviewing the basic setup as well as the graph signal processing and manifold signal processing frameworks.

2.1. Graph signal processing: graph filters

We consider an undirected graph $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with $\mathcal{V}, |\mathcal{V}| = n$, representing the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ the set of edges. The weight function $\mathcal{W} : \mathcal{E} \rightarrow \mathbb{R}$ assigns weights to the edges. Graph signals $\mathbf{x} \in \mathbb{R}^n$ are defined as data supported on the graph, with the i -th element x_i representing the datum at node i [6, 14].

The so-called graph shift operator (GSO) $\mathbf{S} \in \mathbb{R}^{n \times n}$ is typically selected to be the Laplacian matrix, the normalized Laplacian or the adjacency matrix, which encodes the graph sparsity pattern. The ap-

plication of the GSO to a graph signal leads to a diffusion of the data over the edges. Specifically, for each node i , the GSO diffuses the signals on the neighboring nodes x_j scaled with the proximity measure $[\mathbf{S}]_{ij}$ to this node i . I.e. $[\mathbf{S}\mathbf{x}]_i = \sum_{j, (i,j) \in \mathcal{E}} [\mathbf{S}]_{ij} x_j$. Based on this notion of diffusion, we can then define graph convolutional filters as a shift-and-sum operation

$$\mathbf{h}_{\mathbf{G}}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}, \quad (1)$$

where $\{h_k\}_{k=0}^{K-1}$ are the filter coefficients [6, 7].

For an undirected graph \mathbf{G} , the GSO is symmetric and thus can be diagonalized as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$, with $\mathbf{\Lambda}$ containing the graph eigenvalues and \mathbf{V} the graph eigenvectors. By replacing the GSO by its eigendecomposition, we can thus write the spectral representation of the graph convolution as

$$\mathbf{V}^H \mathbf{h}_{\mathbf{G}}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{\Lambda}^k \mathbf{V}^H \mathbf{x} = h(\mathbf{\Lambda}) \mathbf{V}^H \mathbf{x}. \quad (2)$$

The frequency response of graph convolution can therefore be written as $h(\lambda) = \sum_{k=0}^{K-1} h_k \lambda^k$, which only depends on the coefficients h_k and the eigenvalues of \mathbf{S} .

2.2. Manifold signal processing: manifold filters

We consider a compact smooth differentiable d -dimensional submanifold \mathcal{M} embedded in some Euclidean space \mathbb{R}^N , and a measure μ over the manifold \mathcal{M} . Manifold signals are defined as scalar functions $f : \mathcal{M} \rightarrow \mathbb{R}$ supported on \mathcal{M} . The inner product of functions $f, g \in L^2(\mathcal{M})$ can be defined as

$$\langle f, g \rangle_{L^2(\mathcal{M})} = \int_{\mathcal{M}} f(x)g(x)d\mu(x) \quad (3)$$

with $d\mu(x)$ representing the volume element with respect to measure μ . Therefore, the energy of manifold signal f is naturally given as $\|f\|_{L^2(\mathcal{M})}^2 = \langle f, f \rangle_{L^2(\mathcal{M})}$. We consider manifold signals with finite energy, i.e., $f \in L^2(\mathcal{M})$.

The manifold is locally Euclidean and the local homeomorphic Euclidean space around each point $x \in \mathcal{M}$ is defined as the tangent space $T_x \mathcal{M}$. The disjoint union of all tangent spaces over \mathcal{M} is denoted $T\mathcal{M}$. The *intrinsic gradient* $\nabla : L^2(\mathcal{M}) \rightarrow L^2(T\mathcal{M})$ is the differentiation operator and maps scalar functions to tangent vector functions over \mathcal{M} [10]. The adjoint of ∇ is the *intrinsic divergence*, which is defined as $\text{div} : L^2(T\mathcal{M}) \rightarrow L^2(\mathcal{M})$. The Laplace-Beltrami (LB) operator $\mathcal{L} : L^2(\mathcal{M}) \rightarrow L^2(\mathcal{M})$ can be defined as the intrinsic divergence of the intrinsic gradient [18]. More formally,

$$\mathcal{L}f = -\text{div} \circ \nabla f. \quad (4)$$

The LB operator measures the difference between the function value at a point and the average function value over the point's neighborhood [10], similar to graph Laplacian matrix.

The manifold convolution of $\tilde{h} : \mathbb{R}^+ \rightarrow \mathbb{R}$ and the manifold signal $f \in L^2(\mathcal{M})$ is called a manifold filter with impulse response \tilde{h} , and is defined by leveraging the heat diffusion dynamics [8]. Explicitly, this filter, denoted \mathbf{h} , is given by

$$g(x) = (\mathbf{h}f)(x) := \int_0^\infty \tilde{h}(t)e^{-t\mathcal{L}} f(x)dt = \mathbf{h}(\mathcal{L})f(x). \quad (5)$$

This is a spatial manifold convolution operator because it is parametrized by the operator \mathcal{L} and directly operates on points $x \in \mathcal{M}$.

The LB operator \mathcal{L} is self-adjoint and positive-semidefinite. These facts, together with the compactness of \mathcal{M} , imply that \mathcal{L} has a real, positive and discrete eigenvalue spectrum $\{\lambda_i\}_{i=1}^\infty$ such that $\mathcal{L}\phi_i = \lambda_i\phi_i$, where ϕ_i is the eigenfunction associated with eigenvalue λ_i . The eigenvalues are ordered in increasing order as $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots$. The eigenfunctions are orthonormal and thus form an eigenbasis of $L^2(\mathcal{M})$ in the intrinsic sense.

Denoting $[\hat{f}]_i = \langle f, \phi_i \rangle_{L^2(\mathcal{M})} = \int_{\mathcal{M}} f(x)\phi_i(x)d\mu(x)$ and $e^{-t\mathcal{L}}\phi_i = e^{-t\lambda_i}\phi_i$, the manifold convolution can be represented in the manifold frequency domain as

$$[\hat{g}]_i = \int_0^\infty \tilde{h}(t)e^{-t\lambda_i}dt[\hat{f}]_i. \quad (6)$$

Hence, the frequency response of the filter $\mathbf{h}(\mathcal{L})$ is given by $\hat{h}(\lambda) = \int_0^\infty \tilde{h}(t)e^{-t\lambda}dt$, which depends only on the function \tilde{h} and on the eigenvalues of the LB operator. Summing over all i , we thus get

$$g = \mathbf{h}(\mathcal{L})f = \sum_{i=1}^\infty \hat{h}(\lambda_i)[\hat{f}]_i\phi_i. \quad (7)$$

3. CONVERGENCE OF FILTERING ON SAMPLED MANIFOLDS

Our goal is to derive a non-asymptotic approximation bound for manifold filtering on sampled manifolds, providing theoretical guarantees when the sampled manifold is modeled as a graph and showing that graph filtering can be used to approximate manifold filtering. We start by constructing the graph corresponding to the sampled manifold.

3.1. Sampled manifolds as graphs

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of n points sampled uniformly and independently at random from the d -dimensional manifold \mathcal{M} according to measure μ . The empirical measure associated with $d\mu$ is $p_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$, where δ_{x_i} is the Dirac measure supported on x_i . Similar to the definition of inner product in $L^2(\mathcal{M})$ (3), the inner product in $L^2(\mathbf{G}_n)$ is defined as

$$\langle u, v \rangle_{L^2(\mathbf{G}_n)} = \int u(x)v(x)dp_n = \frac{1}{n} \sum_{i=1}^n u(x_i)v(x_i). \quad (8)$$

The norm in $L^2(\mathbf{G}_n)$ with measure p_n is therefore

$$\|u\|_{L^2(\mathbf{G}_n)}^2 = \langle u, u \rangle_{L^2(\mathbf{G}_n)}$$

for $u, v \in L^2(\mathcal{M})$. For signals $\mathbf{u}, \mathbf{v} \in L^2(\mathbf{G}_n)$, the inner product is

$$\langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\mathbf{G}_n)} = \frac{1}{n} \sum_{i=1}^n [\mathbf{u}]_i [\mathbf{v}]_i.$$

From X , we construct an undirected graph \mathbf{G}_n by seeing the sampled points as nodes. The weights of the edges connecting nodes x_i and x_j are determined by

$$w_{ij} = \frac{1}{n} \frac{1}{\epsilon(4\pi\epsilon)^{d/2}} \exp\left(-\frac{\|x_i - x_j\|^2}{4\epsilon}\right), \quad (9)$$

where $\|x_i - x_j\|$ stands for the Euclidean distance between points x_i and x_j and ϵ is a parameter related to the Gaussian kernel [19]. The adjacency matrix $\mathbf{A}_n \in \mathbb{R}^{n \times n}$ can be written as $[\mathbf{A}_n]_{ij} = w_{ij}$ for

$1 \leq i, j \leq n$. Thus, the graph Laplacian associated with this graph can be written as

$$\mathbf{L}_n^\epsilon = \text{diag}(\mathbf{A}_n \mathbf{1}) - \mathbf{A}_n$$

with $[\mathbf{L}_n^\epsilon]_{ij} = -w_{ij}$ if $i \neq j$ and $[\mathbf{L}_n^\epsilon]_{ii} = \sum_{k=1}^n w_{ik}$ [20].

To discretize manifold signals $f \in L^2(\mathcal{M})$ on the sampled manifold, we define a uniform sampling operator $\mathbf{P}_n : L^2(\mathcal{M}) \rightarrow L^2(\mathbf{G}_n)$ yielding signals $\mathbf{f} \in L^2(\mathbf{G}_n)$ defined explicitly as

$$\mathbf{f} = \mathbf{P}_n f \text{ with } \mathbf{f}(x_i) = f(x_i), \quad x_i \in X. \quad (10)$$

This indicates that \mathbf{f} shares the function values of the manifold signal f at the sample points X . Seeing the graph Laplacian as an operator acting on $\mathbf{f} : X \rightarrow \mathbb{R}$, we can write the diffusion operation on each point x explicitly as

$$\mathbf{L}_n^\epsilon \mathbf{f}(x_i) = \sum_{j=1}^n w_{ij} (\mathbf{f}(x_i) - \mathbf{f}(x_j)), \quad i = 1, 2, \dots, n. \quad (11)$$

It can be seen that this formulation can extend to a continuous manifold signal f , which we denote as

$$\mathbf{L}_n^\epsilon f(x) = \frac{1}{n} \frac{1}{\epsilon(4\pi\epsilon)^{d/2}} \sum_{j=1}^n (f(x) - f(x_j)) e^{-\frac{\|x-x_j\|^2}{4\epsilon}}, \quad x \in \mathcal{M}. \quad (12)$$

If we further extend the set of sample points from X to all the points on the manifold, we finally get the functional approximation of the graph Laplacian,

$$\mathbf{L}^\epsilon f(x) = \frac{1}{n} \frac{1}{\epsilon(4\pi\epsilon)^{d/2}} \int_{\mathcal{M}} (f(x) - f(y)) e^{-\frac{\|x-y\|^2}{4\epsilon}} d\mu(y), \quad (13)$$

which operates on all $x \in \mathcal{M}$.

3.2. Graph Laplacian approximation of the LB operator

We first demonstrate the consistency of the discrete graph Laplacians defined in (12) when operating on the eigenfunctions $\phi_i \in L^2(\mathcal{M})$ of the LB operator \mathcal{L} . The quality of the approximation given by the discrete Laplace operator on the sampled manifold is quantified by the following non-asymptotic result.

Theorem 1. Let $\mathcal{M} \subset \mathbb{R}^N$ be a compact smooth differentiable d -dimensional manifold with LB operator \mathcal{L} , whose spectrum is given by $\{\lambda_i, \phi_i\}_{i=1}^\infty$, and assume $\phi_i \in C(\mathcal{M})$. Let \mathbf{G}_n be the discrete graph connecting points $\{x_1, x_2, \dots, x_n\}$ sampled uniformly and independently at random from \mathcal{M} , with edge weights as in (9) and $\epsilon = \epsilon(n) > n^{-1/(d+4)}$. Then, it holds

$$|\mathbf{L}_n^\epsilon \phi_i(x) - \mathcal{L} \phi_i(x)| \leq \left(C_1 \frac{\sqrt{\ln(2/\delta)}}{\sqrt{n\epsilon(4\pi\epsilon)^{d/2}}} + C_2 \sqrt{\epsilon} \right) \lambda_i^{\frac{d+2}{4}}, \quad (14)$$

with probability at least $1 - \delta$. The constants C_1, C_2 depend on the volume of the manifold.

Proof. See Appendix A.1. \square

This theorem provides a point-wise upper bound for the error incurred when using the discrete graph Laplacian to approximate the LB operator operating on the eigenfunctions of \mathcal{L} . We can see that this bound is not only related to the number of sampling points n , but also grows with the corresponding eigenvalue λ_i . This makes sense considering that higher eigenvalues correspond to eigenfunctions with high oscillation [21].

Based on this approximation result, we can further derive a non-asymptotic approximation bound relating the spectra of the graph Laplacian and the LB operator using the Davis-Khan theorem [22]. This result, stated in Theorem 2, will allow us to analyze the approximation of a manifold filter by a graph filter through their spectral representations.

Theorem 2. Let $\mathcal{M} \subset \mathbb{R}^N$ be a compact smooth differentiable d -dimensional manifold with LB operator \mathcal{L} whose spectrum is given by $\{\lambda_i, \phi_i\}_{i=1}^\infty$. Let \mathbf{L}_n^ϵ be a discrete graph Laplacian defined as in (12) with $\{x_1, x_2, \dots, x_n\}$ sampled uniformly and independently at random from \mathcal{M} whose spectrum is given by $\{\lambda_{i,n}^\epsilon, \phi_{i,n}^\epsilon\}_{i=1}^n$. Fix $K \in \mathbb{N}$ and assume that n is sufficiently large so that $\epsilon = \epsilon(n) > n^{-1/(d+4)}$. Then, with probability at least $1 - 2e^{-n}$, we have

$$|\lambda_i - \lambda_{i,n}^\epsilon| \leq \Omega_1 \sqrt{\epsilon}, \quad \|a_i \phi_{i,n}^\epsilon - \phi_i\| \leq \Omega_2 \sqrt{\epsilon}, \quad (15)$$

with $a_i = \{-1, 1\}$ for all $i < K$. The constants Ω_1, Ω_2 depend on λ_K , the eigengap of \mathcal{L} i.e., $\theta = \min_{1 \leq i \leq K} \{\lambda_i - \lambda_{i-1}, \lambda_{i+1} - \lambda_i\}$, d and the volume of \mathcal{M} .

Proof. See Appendix A.2. \square

3.3. Graph filtering approximation of manifold filtering

Equipped with theoretical approximation results for both the eigenvalues and eigenfunctions of the Laplacian operators, we can now prove that graph filters can approximate manifold filters well in the spectral domain. We first show that we can generalize the definition of manifold convolutional filters to sampled manifolds using the same impulse response \tilde{h} .

From the definition in (5), recall that the manifold filter is parametric on the LB operator when the impulse response $\tilde{h}(t)$ is fixed. By replacing the LB operator with the discrete graph Laplacian defined in (12), we can then obtain a graph convolutional filter

$$\mathbf{g} = \int_0^\infty \tilde{h}(t) e^{-t\mathbf{L}_n^\epsilon} \mathbf{f} dt = \mathbf{h}(\mathbf{L}_n^\epsilon) \mathbf{f}, \quad \mathbf{g}, \mathbf{f} \in \mathbb{R}^n. \quad (16)$$

This can be seen as a *continuous-time* graph filtering process, different from the *discrete-time* graph filtering process \mathbf{h}_G in (1), where the exponential term $e^{-\mathbf{L}_n^\epsilon}$ can be seen as the graph shift operator. The frequency representation of this graph filter can be written as

$$\mathbf{g} = \sum_{i=1}^n \hat{h}(\lambda_{i,n}^\epsilon) \langle \mathbf{f}, \phi_{i,n}^\epsilon \rangle_{L^2(\mathbf{G}_n)} \phi_{i,n}^\epsilon, \quad (17)$$

which reveals the full dependency on the eigendecomposition of \mathbf{L}_n^ϵ . Thus, we can relate graph filtering and manifold filtering using the spectral approximation results for the graph Laplacian and LB operator presented in Theorem 2.

Unlike the finite spectrum of the graph Laplacian, the LB operator \mathcal{L} possesses an infinite spectrum. We consider Weyl's law [23] when analyzing the spectral properties of \mathcal{L} . This classical result states that the eigenvalues λ_i of \mathcal{L} grow in the order of $i^{2/d}$. Therefore, the difference between neighboring eigenvalues becomes quite small in the high frequency spectrum, i.e., large eigenvalues accumulate. Leveraging this fact, we can use a partition strategy to separate the spectrum into finite groups as in Definition 1.

Definition 1. [13] (α -separated spectrum) The α -separated spectrum of LB operator \mathcal{L} is defined as the partition $\Lambda_1(\alpha) \cup \dots \cup \Lambda_N(\alpha)$ such that $\lambda_i \in \Lambda_k(\alpha)$ and $\lambda_j \in \Lambda_l(\alpha)$, $k \neq l$, satisfying $|\lambda_i - \lambda_j| > \alpha$.

The following defined α -FDT filter allows obtaining the α -separated spectrum as in Definition 1.

Definition 2. [13] (α -FDT filter) The α -frequency difference threshold (α -FDT) filter is defined as a filter $\mathbf{h}(\mathcal{L})$ whose frequency response satisfies

$$|\hat{h}(\lambda_i) - \hat{h}(\lambda_j)| \leq \gamma_k \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\alpha) \quad (18)$$

with $\gamma_k \leq \gamma$ for $k = 1, \dots, N$.

To derive an approximation bound, we will also need the manifold filter to have a Lipschitz frequency response as in Definition 3.

Definition 3 (Lipschitz filter). A filter is A_h -Lipschitz if its frequency response is Lipschitz continuous with Lipschitz constant A_h ,

$$|\hat{h}(a) - \hat{h}(b)| \leq A_h |a - b| \text{ for all } a, b \in (0, \infty). \quad (19)$$

Equipped with the above requirements for the filter frequency response, we can finally establish the upper bound on the manifold filter approximation error on the sampled manifold.

Theorem 3. Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of n points sampled by an operator \mathbf{P}_n (10) from a d -dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$. Let \mathbf{G}_n be a discrete graph approximation of \mathcal{M} constructed from X with weight values as in (9) with $\epsilon = \epsilon(n) \geq n^{-1/(d+4)}$. Let $\mathbf{h}(\cdot)$ be the convolutional filter parameterized by the LB operator \mathcal{L} of the manifold \mathcal{M} (5) or by the discrete graph Laplacian operator \mathbf{L}_n^ϵ of the graph \mathbf{G}_n . Under the assumption that the frequency response of filter \mathbf{h} is Lipschitz continuous and α -FDT with $\alpha^2 \gg \epsilon$ and $\gamma = \Omega_2' \sqrt{\epsilon}/\alpha$, with probability at least $1 - 2n^{-2}$ it holds that

$$\begin{aligned} & \|\mathbf{h}(\mathbf{L}_n^\epsilon) \mathbf{P}_n f - \mathbf{P}_n \mathbf{h}(\mathcal{L}) f\|_{L^2(\mathbf{G}_n)} \\ & \leq \left(\frac{N\Omega_2'}{\alpha} + A_h \Omega_1 \right) \sqrt{\epsilon} + C_{ge} \sqrt{\frac{\log n}{n}} \end{aligned} \quad (20)$$

where N is the partition size of α -FDT filter and C_{ge} is related with d and the volume of \mathcal{M} .

Proof. See Appendix A.3. \square

From this theorem, we see that if we take $\epsilon = n^{-1/(d+4)}$, the difference between filtering on the manifold and the sampled manifold is in the order of $O(n^{-1/(2d+8)})$. Thus, with a large enough number of sampling points, this approximation provides a good accuracy with high probability. This provides a formal theoretical guarantee for the ability of graph filters to approximate manifold filters on sampled manifolds, enabling approximate linear information processing on continuous non-Euclidean domains.

4. NUMERICAL EXPERIMENTS

We consider a problem of automatic navigation control of an agent. We intend to control an agent to find a path to the goal without colliding to obstacles. We sample grid points over the free space which means avoiding the obstacle areas to construct a graph structure. We generate four trajectories with Dijkstra's shortest path algorithm leading a starting point to the goal. Each point along the trajectory is labeled with the direction of the velocity. The adjacency matrix can be calculated based on the weight value defined in (9) with the Euclidean distance between each two points. We note that if the direct path between two points goes through the obstacles, we set the distance between them as infinity, i.e. the weight value as zero, so as to better capture the geometric structure. The input graph signal is the position of the nodes and the output of the graph filter is the direction of this node leading to the goal. We construct a one-layer and two-layer graph filters with 5 filter taps in each layer. We train all the

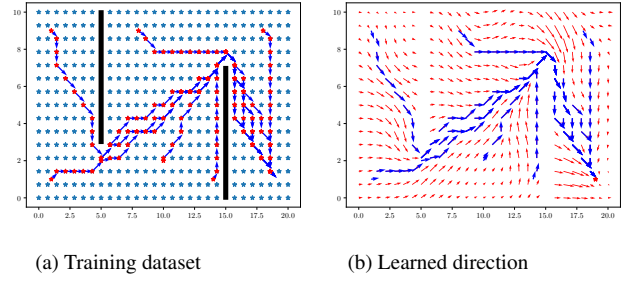


Fig. 1: Graph filtering performance with $n = 413$

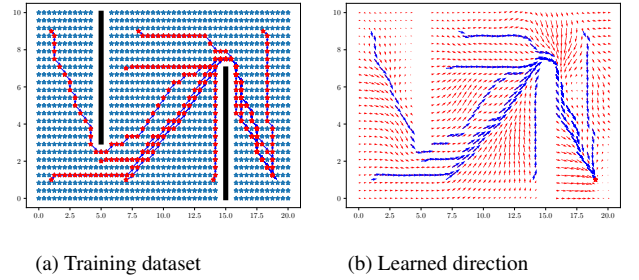


Fig. 2: Graph filtering performance with $n = 1117$

architectures for 30,000 epochs with SGD optimizer with the learning rate set as 0.0002. The performance of the learned graph filter is testified by randomly generating 100 starting points and computing the trajectories. If the trajectory can reach the goal without colliding into the obstacles, the trajectory is marked as “success”.

We first sample n points uniformly over the free space avoiding the black obstacles as shown in Fig. 1a ($n = 413$) and Fig. 2a ($n = 1117$). In Fig. 1a and Fig. 2a, the red stars depicting labeled points as training dataset and the blue arrows attached to the red stars point to the optimal directions leading to the goal. The blue stars depict unlabeled points over the free space. We implement a 1-layer or 2-layer graph filters to learn and estimate the directions for all the unlabeled points. Figure 1b and 2b show the learned directions starting from each unlabeled points as the red arrows show using graph filtering. The numbers of successful trajectories within 100 generated testing are shown in Table 1. We can see that graph filtering can efficiently learn the potential successful trajectories for unlabeled points based on these sampled manifolds. More sampling points characterize the free space more accurately, which leads to more successful testing trajectories. This is also in accordance with our theoretical result presented in Theorem 3. Moreover, more layers of filters also learn a more accurate prediction.

	$n = 413$	$n = 1117$
1-layer graph filter	74	75
2-layer graph filter	79	84

Table 1: Number of successful trajectories.

5. CONCLUSIONS

In this paper, we introduced manifold convolutional filters defined with the exponential Laplace-Beltrami operator to process geometric data. To access the manifold model, we sample uniformly over the

manifold and construct a graph model as a sampled manifold. We first prove the non-asymptotic approximation error bound of discrete graph Laplacian to LB operator in both operator and spectral aspects. By transferring the manifold filtering to the sampled manifold, the approximation of the graph filtering to the manifold filtering can be derived. The non-asymptotic error bound decreases with the number of sampling points growing. The approximation of graph filtering is verified empirically on a navigation control problem over a manifold.

6. REFERENCES

- [1] Q. Li, F. Gama, A. Ribeiro, and A. Prorok, "Graph neural networks for decentralized multi-robot path planning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 785–11 792.
- [2] Z. Wang and M. Gombolay, "Learning scheduling policies for multi-robot coordination with graph attention networks," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4509–4516, 2020.
- [3] Z. Wang, L. Ruiz, M. Eisen, and A. Ribeiro, "Stable and transferable wireless resource allocation policies via manifold neural networks," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 8912–8916.
- [4] J. Zeng, G. Cheung, M. Ng, J. Pang, and C. Yang, "3d point cloud denoising using graph laplacian regularization of a low dimensional manifold model," *IEEE Transactions on Image Processing*, vol. 29, pp. 3474–3489, 2019.
- [5] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo, "3-d human action recognition by shape analysis of motion trajectories on riemannian manifold," *IEEE transactions on cybernetics*, vol. 45, no. 7, pp. 1340–1352, 2014.
- [6] 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.
- [7] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro, "Convolutional neural network architectures for signals supported on graphs," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1034–1049, 2019.
- [8] Z. Wang, L. Ruiz, and A. Ribeiro, "Convolutional neural networks on manifolds: From graphs and back," *arXiv preprint arXiv:2210.00376*, 2022.
- [9] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst, "Geodesic convolutional neural networks on riemannian manifolds," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 37–45.
- [10] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.
- [11] W. Shi and R. Rajkumar, "Point-gnn: Graph neural network for 3d object detection in a point cloud," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1711–1719.
- [12] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model cnns," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5115–5124.
- [13] Z. Wang, L. Ruiz, and A. Ribeiro, "Stability of neural networks on riemannian manifolds," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1845–1849.
- [14] 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*, vol. 30, no. 3, pp. 83–98, 2013.
- [15] L. Ruiz, L. Chamon, and A. Ribeiro, "Graphon neural networks and the transferability of graph neural networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1702–1712, 2020.
- [16] N. Keriven, A. Bietti, and S. Vaiter, "Convergence and stability of graph convolutional networks on large random graphs," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 512–21 523, 2020.
- [17] J. Calder and N. G. Trillos, "Improved spectral convergence rates for graph laplacians on epsilon-graphs and k-nn graphs," *arXiv preprint arXiv:1910.13476*, 2019.
- [18] S. Rosenberg and R. Steven, *The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds*. Cambridge University Press, 1997, no. 31.
- [19] M. Belkin and P. Niyogi, "Towards a theoretical foundation for laplacian-based manifold methods," *Journal of Computer and System Sciences*, vol. 74, no. 8, pp. 1289–1308, 2008.
- [20] R. Merris, "A survey of graph laplacians," *Linear and Multilinear Algebra*, vol. 39, no. 1-2, pp. 19–31, 1995.
- [21] Y. Shi and B. Xu, "Gradient estimate of an eigenfunction on a compact riemannian manifold without boundary," *Annals of Global Analysis and Geometry*, vol. 38, no. 1, pp. 21–26, 2010.
- [22] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the davis–kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2015.
- [23] W. Arendt, R. Nittka, W. Peter, and F. Steiner, "Weyl's law: Spectral properties of the laplacian in mathematics and physics," *Mathematical analysis of evolution, information, and complexity*, pp. 1–71, 2009.
- [24] M. Belkin and P. Niyogi, "Convergence of laplacian eigenmaps," *Advances in neural information processing systems*, vol. 19, 2006.
- [25] U. Von Luxburg, M. Belkin, and O. Bousquet, "Consistency of spectral clustering," *The Annals of Statistics*, pp. 555–586, 2008.
- [26] D. B. Dunson, H.-T. Wu, and N. Wu, "Spectral convergence of graph laplacian and heat kernel reconstruction in 1 from random samples," *Applied and Computational Harmonic Analysis*, vol. 55, pp. 282–336, 2021.

A. APPENDIX

A.1. Proof of Theorem 1

We decompose the operator difference between the graph Laplacian and the LB operator with an intermediate term \mathbf{L}^ϵ , which is the functional approximation defined in (13). We first focus on the operator difference between \mathbf{L}^ϵ and \mathcal{L} . From [19], we can get the bound as

$$\|\mathbf{L}^\epsilon \phi_i - \mathcal{L} \phi_i\| \leq C \sqrt{\epsilon} \|\phi_i\|_{H^{d/2+1}}, \quad (21)$$

For the Sobolev norm of eigenfunction ϕ_i , according to [24, Lemma 4.4] we have

$$\|\phi_i\|_{H^{d/2+1}} \leq C\lambda_i^{\frac{d+2}{4}}, \quad (22)$$

which leads to

$$\|\mathbf{L}^\epsilon \phi_i - \mathcal{L} \phi_i\| \leq C_1 \sqrt{\epsilon} \lambda_i^{\frac{d+2}{4}}. \quad (23)$$

For the operator difference between \mathbf{L}_n^ϵ and \mathbf{L}^ϵ with Hoeffding's inequality as

$$\mathbb{P}(|\mathbf{L}_n^\epsilon \phi_i(x) - \mathbf{L}^\epsilon \phi_i(x)| > \epsilon_1) \leq \exp\left(-\frac{2n\epsilon_1^2}{\|\phi_i\|_{H^{d/2+1}}^2}\right). \quad (24)$$

Therefore, we can claim that with probability at least $1 - \delta$, we have

$$|\mathbf{L}_n^\epsilon \phi_i(x) - \mathbf{L}^\epsilon \phi_i(x)| \leq \sqrt{\frac{\ln 1/\delta}{2n}} \|\phi_i\|_{H^{d/2+1}}. \quad (25)$$

Combining (23) and (25) with triangle inequality, we can get the conclusion in Theorem 1.

A.2. Proof of Theorem 2

We first import two lemmas to help prove the spectral properties.

Lemma 1. Let \mathbf{A}, \mathbf{B} be self-adjoint operators with $\{\lambda_i(\mathbf{A}), \mathbf{u}_i\}_{i=1}^\infty$ and $\{\lambda_i(\mathbf{B}), \mathbf{w}_i\}_{i=1}^\infty$ as the corresponding spectrum. Let $Pr_{\mathbf{w}_i}$ be the orthogonal projection operation onto the subspace generated by \mathbf{w}_i . Then we have

$$\|a_i \mathbf{u}_i - \mathbf{w}_i\| \leq 2\|\mathbf{u}_i - Pr_{\mathbf{w}_i} \mathbf{u}_i\| \leq \frac{2\|\mathbf{B}\mathbf{u}_i - \mathbf{A}\mathbf{u}_i\|}{\min_{j \neq i} |\lambda_j(\mathbf{B}) - \lambda_i(\mathbf{A})|}. \quad (26)$$

Proof. The first inequality is directly from [25, Proposition 18]. Let $Pr_{\mathbf{w}_i}^\perp$ be the orthogonal projection onto the complement of the subspace generated by \mathbf{w}_i . Then we have

$$\|\mathbf{u}_i - Pr_{\mathbf{w}_i} \mathbf{u}_i\| = \|Pr_{\mathbf{w}_i}^\perp \mathbf{u}_i\| = \left\| \sum_{j \neq i} \langle \mathbf{u}_i, \mathbf{w}_j \rangle \mathbf{w}_j \right\|. \quad (27)$$

Therefore, we have

$$\begin{aligned} & \|Pr_{\mathbf{w}_i}^\perp \mathbf{B}\mathbf{u}_i - Pr_{\mathbf{w}_i}^\perp \mathbf{A}\mathbf{u}_i\| \\ &= \left\| \sum_{j \neq i} \langle \mathbf{B}\mathbf{u}_i, \mathbf{w}_j \rangle \mathbf{w}_j - \sum_{j \neq i} \langle \mathbf{A}\mathbf{u}_i, \mathbf{w}_j \rangle \mathbf{w}_j \right\| \end{aligned} \quad (28)$$

$$= \left\| \langle \mathbf{u}_i, \mathbf{B}\mathbf{w}_j \rangle \mathbf{w}_j - \sum_{j \neq i} \lambda_i(\mathbf{A}) \langle \mathbf{u}_i, \mathbf{w}_j \rangle \mathbf{w}_j \right\| \quad (29)$$

$$= \left\| \sum_{j \neq i} (\lambda_i(\mathbf{B}) - \lambda_i(\mathbf{A})) \langle \mathbf{u}_i, \mathbf{w}_j \rangle \mathbf{w}_j \right\| \quad (30)$$

$$\geq \min_{j \neq i} |\lambda_i(\mathbf{B}) - \lambda_i(\mathbf{A})| \left\| \sum_{j \neq i} \langle \mathbf{u}_i, \mathbf{w}_j \rangle \mathbf{w}_j \right\| \quad (31)$$

$$= \min_{j \neq i} |\lambda_i(\mathbf{B}) - \lambda_i(\mathbf{A})| \|\mathbf{u}_i - Pr_{\mathbf{w}_i} \mathbf{u}_i\|, \quad (32)$$

together with $\|\mathbf{B}\mathbf{u}_i - \mathbf{A}\mathbf{u}_i\| \geq \|Pr_{\mathbf{w}_i}^\perp \mathbf{B}\mathbf{u}_i - Pr_{\mathbf{w}_i}^\perp \mathbf{A}\mathbf{u}_i\|$. We can conclude the proof. \square

The following lemma is adapted from [26, Lemma 5c]

Lemma 2. Let \mathbf{A}, \mathbf{B} be self-adjoint operators with $\{\lambda_i(\mathbf{A}), \mathbf{u}_i\}_{i=1}^\infty$ and $\{\lambda_i(\mathbf{B}), \mathbf{w}_i\}_{i=1}^\infty$ as the corresponding spectrum. Then we have

$$|\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B})| = \frac{\langle (\mathbf{A} - \mathbf{B})\mathbf{u}_i, \mathbf{w}_i \rangle}{|\langle \mathbf{u}_i, \mathbf{w}_i \rangle|} \leq \frac{\|(\mathbf{A} - \mathbf{B})\mathbf{u}_i\|}{|\langle \mathbf{u}_i, \mathbf{w}_i \rangle|} \quad (33)$$

With the above lemmas and our proposed Theorem 1, which includes the operator difference, we can prove Theorem 2. We first fix some $K \in \mathbb{N}$, which provides an upper bound for $\lambda_i \leq \lambda_K$ for all $1 \leq i \leq K$. By taking the probability $1 - n^{-2}$ and $\epsilon = n^{-1/(d+4)}$, we can conclude that the operator difference in Theorem 1 can be bounded with order $O(\sqrt{\epsilon})$, with the constant scaling with $\lambda_K^{\frac{d+2}{4}}$. Combine with Lemma 1 and $\theta = \min_{1 \leq j \neq i \leq K} |\lambda_j - \lambda_i|$, we can get

$$\|a_i \phi_{i,n}^\epsilon - \phi_i\| \leq \frac{C_k}{\theta} \sqrt{\epsilon}, \quad (34)$$

where we denote the constant as Ω_1 to include the effects of K , the eigengap and the volume of \mathcal{M} .

This upper bound of the eigenfunction difference leads to $|\langle \mathbf{u}_i, \mathbf{w}_i \rangle| \geq 1 - \Omega_1/2\sqrt{\epsilon} \geq 1$. Combining with Lemma 2, the difference of the eigenvalues can also be bounded in the order of $O(\sqrt{\epsilon})$.

A.3. Proof of Theorem 3

We first write out the filter representation as

$$\begin{aligned} & \|\mathbf{h}(\mathbf{L}_n^\epsilon) \mathbf{P}_n f - \mathbf{P}_n \mathbf{h}(\mathcal{L}) f\| \\ & \leq \left\| \sum_{i=1}^\infty \hat{h}(\lambda_{i,n}^\epsilon) \langle \mathbf{P}_n f, \phi_{i,n}^\epsilon \rangle_{\mathbf{G}_n} \phi_i^n - \sum_{i=1}^\infty \hat{h}(\lambda_i) \langle f, \phi_i \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i \right\| \end{aligned} \quad (35)$$

We decompose the α -FDT filter function as $\hat{h}(\lambda) = h^{(0)}(\lambda) + \sum_{l \in \mathcal{K}_m} h^{(l)}(\lambda)$ as

$$h^{(0)}(\lambda) = \begin{cases} \hat{h}(\lambda) - \sum_{l \in \mathcal{K}_m} \hat{h}(C_l) & \lambda \in [\Lambda_k(\alpha)]_{k \in \mathcal{K}_s} \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

$$h^{(l)}(\lambda) = \begin{cases} \hat{h}(C_l) & \lambda \in [\Lambda_k(\alpha)]_{k \in \mathcal{K}_s} \\ \hat{h}(\lambda) & \lambda \in \Lambda_l(\alpha) \\ 0 & \text{otherwise} \end{cases} \quad (37)$$

With the triangle inequality, we start by analyzing the output difference of $h^{(0)}(\lambda)$ as

$$\begin{aligned} & \left\| \sum_{i=1}^\infty h^{(0)}(\lambda_i^n) \langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n - \sum_{i=1}^\infty h^{(0)}(\lambda_i) \langle f, \phi_i \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i \right\| \\ & \leq \left\| \sum_{i=1}^\infty (h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i)) \langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n \right\| \\ & + \left\| \sum_{i=1}^\infty h^{(0)}(\lambda_i) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n - \langle f, \phi_i \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i) \right\|. \end{aligned} \quad (38)$$

The first term in (38) can be bounded by leveraging the A_h -Lipschitz continuity of the frequency response. From the eigenvalue difference in Theorem 2, we can claim that for each eigenvalue $\lambda_i \leq \lambda_K$, we have

$$|\lambda_{i,n}^\epsilon - \lambda_i| \leq \sqrt{\epsilon}. \quad (39)$$

The first term is bounded as

$$\begin{aligned} & \left\| \sum_{i=1}^\infty (h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i)) \langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n \right\| \\ & \leq \sum_{i=1}^\infty |h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i)| \|\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n}\| \|\phi_i^n\| \end{aligned} \quad (40)$$

$$\leq \sum_{i=1}^{N_s} A_h |\lambda_i^n - \lambda_i| \|\mathbf{P}_n f\| \|\phi_i^n\|^2 \leq N_s A_h \Omega_1 \sqrt{\epsilon}. \quad (41)$$

The second term in (38) can be bounded combined with the convergence of eigenfunctions in (43) as

$$\begin{aligned} & \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n - \langle f, \phi_i \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i) \right\| \\ & \leq \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n - \langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \mathbf{P}_n \phi_i) \right\| \\ & + \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \mathbf{P}_n \phi_i - \langle f, \phi_i \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i) \right\| \quad (42) \end{aligned}$$

From the convergence stated in Theorem 2, we have

$$\mathbb{P}(|\phi_i^n(x_j) - \phi_i(x_j)| \leq \epsilon_i) \geq 1 - \delta_i, \quad \text{for all } x_j \in X. \quad (43)$$

Therefore, letting $\epsilon_i < \epsilon$ with $\epsilon > 0$, with probability at least $\prod_{i=1}^M (1 - \delta_i) := 1 - \delta$, for all $n > \max_i N_i := N$, the first term in (42) can be bounded as

$$\begin{aligned} & \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \phi_i^n - \langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i) \right\| \\ & \leq \sum_{i=1}^{N_s} \|\mathbf{P}_n f\| \|\phi_i^n - \mathbf{P}_n \phi_i\| \leq N_s \Omega_2 \sqrt{\epsilon}. \quad (44) \end{aligned}$$

The last equation comes from the definition of norm in $L^2(\mathbf{G}_n)$. The second term in (42) can be written as

$$\begin{aligned} & \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i^n) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} \mathbf{P}_n \phi_i - \langle f, \phi_i \rangle_{\mathcal{M}} \mathbf{P}_n \phi_i) \right\| \\ & \leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i^n)| |\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} - \langle f, \phi_i \rangle_{\mathcal{M}}| \|\mathbf{P}_n \phi_i\|. \quad (45) \end{aligned}$$

Because $\{x_1, x_2, \dots, x_n\}$ is a set of uniform sampled points from \mathcal{M} , based on Theorem 19 in [25] we can claim that

$$|\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} - \langle f, \phi_i \rangle_{\mathcal{M}}| = O\left(\sqrt{\frac{\log n}{n}}\right). \quad (46)$$

Taking into consider the boundedness of frequency response $|h^{(0)}(\lambda)| \leq 1$ and the bounded energy $\|\mathbf{P}_n \phi_i\|$. Therefore, we have

$$\left\| \sum_{i=1}^N \hat{h}(\lambda_i^n) (\langle \mathbf{P}_n f, \phi_i^n \rangle_{\mathbf{G}_n} - \langle f, \phi_i \rangle_{\mathcal{M}}) \mathbf{P}_n \phi_i \right\| = O\left(\sqrt{\frac{\log n}{n}}\right).$$

Combining the above results, we can bound the output difference of $h^{(0)}$. Then we need to analyze the output difference of $h^{(l)}(\lambda)$ and bound this as

$$\begin{aligned} & \left\| \mathbf{P}_n \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathbf{L}_n) \mathbf{P}_n f \right\| \\ & \leq \left\| (\hat{h}(C_l) + \delta) \mathbf{P}_n f - (\hat{h}(C_l) - \delta) \mathbf{P}_n f \right\| \leq 2\delta \|\mathbf{P}_n f\|, \quad (47) \end{aligned}$$

where $\mathbf{h}^{(l)}(\mathcal{L})$ and $\mathbf{h}^{(l)}(\mathbf{L}_n)$ are filters with filter function $h^{(l)}(\lambda)$ on the LB operator \mathcal{L} and graph Laplacian \mathbf{L}_n respectively. Combining

the filter functions, we can write

$$\begin{aligned} & \|\mathbf{P}_n \mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathbf{L}_n) \mathbf{P}_n f\| \\ & = \left\| \mathbf{P}_n \mathbf{h}^{(0)}(\mathcal{L})f + \mathbf{P}_n \sum_{l \in \mathcal{K}_m} \mathbf{h}^{(l)}(\mathcal{L})f - \right. \\ & \quad \left. \mathbf{h}^{(0)}(\mathbf{L}_n) \mathbf{P}_n f - \sum_{l \in \mathcal{K}_m} \mathbf{h}^{(l)}(\mathbf{L}_n) \mathbf{P}_n f \right\| \quad (48) \end{aligned}$$

$$\begin{aligned} & \leq \|\mathbf{P}_n \mathbf{h}^{(0)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathbf{L}_n) \mathbf{P}_n f\| + \\ & \quad \sum_{l \in \mathcal{K}_m} \|\mathbf{P}_n \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathbf{L}_n) \mathbf{P}_n f\|. \quad (49) \end{aligned}$$

Above all, we can claim that

$$\|\mathbf{h}(\mathbf{L}_n) \mathbf{P}_n f - \mathbf{P}_n \mathbf{h}(\mathcal{L})f\| \leq (N\Omega_2 + A_h\Omega_1) \sqrt{\epsilon} + C \sqrt{\frac{\log n}{n}} \quad (50)$$