Stability to Deformations of Manifold Filters and Manifold Neural Networks

Zhiyang Wang   Luana Ruiz   Alejandro Ribeiro

Abstract—The paper defines and studies manifold (M) convolutional filters and neural networks (NNs). *Manifold* filters and MNNs are defined in terms of the Laplace-Beltrami operator exponential and are such that *graph* (G) filters and neural networks (NNs) are recovered as discrete approximations when the manifold is sampled. These filters admit a spectral representation which is a generalization of both the spectral representation of graph filters and the frequency response of standard convolutional filters in continuous time. The main technical contribution of the paper is to analyze the stability of manifold filters and MNNs to smooth deformations of the manifold. This analysis generalizes known stability properties of graph filters and GNNs and it is also a generalization of known stability properties of standard convolutional filters and neural networks in continuous time. The most important observation that follows from this analysis is that manifold filters, same as graph filters and standard continuous time filters, have difficulty discriminating high frequency components in the presence of deformations. This is a challenge that can be ameliorated with the use of manifold, graph, or continuous time neural networks. The most important practical consequence of this analysis is to shed light on the behavior of graph filters and GNNs in large scale graphs.

Index Terms—Graph Signal Processing, Graph Neural Networks, Manifolds, Manifold Filters, Manifold Neural Networks, Manifold Deformations, Operator Stability.

I. INTRODUCTION

Graph convolutional filters [3]–[5] and graph neural networks (GNNs) [6]–[8] have become the tool of choice for signal and information processing on graphs, e.g., [9]–[12]. In several applications, graphs can be considered as samples of a manifold. This is sometimes quite explicit as in the case of, e.g., point clouds [13], [14] and sometimes somewhat implicit as in the case of, e.g., wireless communication networks [9], [10]. In this context we can think of convolutions on graphs and GNNs as discretizations of their manifold counterparts. This is important because one can often gain valuable insights about discrete (graph) signal processing by studying continuous (manifold) signal processing.

The main technical contribution of this paper is analyzing the stability of manifold filters and manifold neural networks (MNNs) to smooth deformations of the manifold. Prior to developing stability analyses we define *manifold* convolutional filters and frequency responses that are consistent with *graph* convolutional filters and frequency responses.

Manifold Filters. Consider a manifold with Laplace-Beltrami (LB) operator $\mathcal{L}$ and formulate the corresponding manifold diffusion equation with respect to an auxiliary time variable t. For initial condition $f(x)$, manifold diffusions generate the time varying manifold function $u(x,t) = e^{-t\mathcal{L}}f(x)$ in which $e^{-t\mathcal{L}}$ is the LB operator exponential (Section II-A). A manifold convolutional filter $\tilde{h}(t)$ then acts on the function $f(x)$ as the integral over time of the product of $\tilde{h}(t)$ with the diffusion sequence $u(x,t)$ (Section II-B),

$$g(x) = (\mathbf{h}f)(x) = \int_0^\infty \tilde{h}(t)e^{-t\mathcal{L}}f(x)\mathrm{d}t. \tag{1}$$

A manifold convolutional filter is such that we can recover graph filters through discretization of the manifold and discretization of the auxiliary time variable t (Section VI). The reason for this connection is that graph filters are linear combinations of the elements of the graph diffusion sequence [3], [15], [16] and (1) defines manifold filters as linear combinations of the elements of the manifold diffusion sequence. Manifold filters are also generalizations of standard time convolutions. This requires consideration of the wave equation on the real line so that the exponential of the derivative operator $e^{-t\partial/\partial x}$ appears in (1) (Appendix 1).

Frequency Response of Manifold Filters. We define a filter's frequency response as

$$\hat{h}(\lambda) = \int_0^\infty \tilde{h}(t)e^{-t\lambda}\mathrm{d}t. \tag{2}$$

This definition is motivated by the fact that manifold convolutions can be decomposed on separate spectral components (Section II-C). Indeed, let ϕ_i be an eigenfunction of the LB operator associated with eigenvalue λ_i. If $[\hat{f}]_i = \langle f, \phi_i \rangle_{L^2(\mathcal{M})}$ and $[\hat{g}]_i = \langle g, \phi_i \rangle_{L^2(\mathcal{M})}$ are projections of the functions f and g in (1) on this eigenfunction, $[\hat{g}]_i$ depends only on $[\hat{f}]_i$ and can be written as $[\hat{g}]_i = \hat{h}(\lambda_i)[\hat{f}]_i$ (Proposition 2).

This definition of the frequency response of a manifold filter generalizes the frequency response of a standard time filter. This is because the definition in (2) is a Laplace transform, which reduces to the Fourier transform of a filter's impulse response when restricted to the imaginary axis $\lambda = j\omega$. It is also a generalization of the frequency response of a graph filter, which is a z-transform [17].

Stability to Deformations. We establish stability of manifold filters and MNNs with respect to domain deformations formally defined as manifold diffeomorphisms. We consider a signal f defined on a manifold $\mathcal{M}$ and the signal $f \circ \tau$ made up of the composition of f with a diffeomorphism. The signals f and $f \circ \tau$ are passed through the same MNN $\boldsymbol{\Phi}$. In this paper we

prove that if τ is ϵ-small and ϵ-smooth the respective MNN outputs satisfy

$$\|\boldsymbol{\Phi}(f \circ \tau) - \boldsymbol{\Phi}(f)\|_{L^2(\mathcal{M})} = O(\epsilon)\|f\|_{L^2(\mathcal{M})}, \qquad (3)$$

provided that the manifold filters in the layers of the MNN satisfy certain spectral properties (Sections III and IV).

The bound in (3) is a generalization of the standard convolutional neural network (CNN) bound in [18], which studies diffeomorphisms of the real line and its effect on convolutional filter banks and CNNs. The bound is also a limit version of the GNN bounds of [19] and related literature [20]–[24]. As is the case of [18]–[24] the bound in (3) holds when the frequency response $\hat{h}(\lambda)$ of the filters that make up the layers of the MNN have decreasing variability with increasing λ (Section III). Thus, stability requires *layers* with limited ability to discriminate high frequency components and implies that one may expect *multi-layered* MNNs to outperform filters in learning tasks in which high frequency components are important (Section V).

Related Work and Significance. We focus on the stability analysis of MNNs – the limit version of GNNs – because all of the existing GNN stability results have bounds that grow with the number of nodes [20], [23] in the graph. To overcome this limitation, many works have studied neural networks on graphons [24]–[26] and more general graph models with variable sparsity [27]. Results in these settings are independent of graph size – same as the results for a limit object presented here. It needs to be pointed out that graphons are limits of dense graphs in the sense that they have growing degrees. This is different from graphs sampled from manifolds that can have finite degrees as the discretization becomes finer [28]. Moreover, even if there exist random graph models allowing to model moderately sparse graphs (e.g., [27]), these models do not have the physical interpretation of a manifold, which is often a better descriptor of real-world domains. Of particular relevance to our paper is the work on GNN transferability for graphs that are sampled from a general topological space [29].

Organization. Section II introduces preliminary definitions (Section II-A), defines manifold convolutions (Section II-B), and introduces the spectral domain representation of manifold filters (Section II-C). Section III studies the stability of manifold filters to manifold deformations. It shows that a diffeomorphism results in a perturbation of the LB operator that involves additive and multiplicative terms (Theorem 1). It then goes on to study the effect on manifold filters of additive (Section III-A) and multiplicative (Section III-B) perturbations of the LB operator. Section IV extends the analysis of manifold filter stability to manifold neural networks and Section V discusses the implications of the results derived in Sections III and IV. Section VI explains how to recover graph filters and GNNs from the discretization of a manifold. Section VII illustrates the results of Sections III and IV with numerical examples. Section VIII concludes the paper. Proofs are deferred to appendices and supplmenetary material.

II. MANIFOLD CONVOLUTIONAL FILTERS

Consider a compact, smooth and differentiable d-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^N$. For simplicity, in this paper we use the words submanifold and manifold interchangeably, assuming the manifold $\mathcal{M}$ to always be embedded in $\mathbb{R}^N$. This embedding induces a Riemannian structure [30]. In turn, the Riemannian structure allows defining a measure μ over the manifold as well as a notion of *length* for smooth curves on $\mathcal{M}$. Given two points $x, y \in \mathcal{M}$, the length of the shortest curve between x and y is denoted dist(x,y) and called the geodesic distance between these points.

We consider the manifold $\mathcal{M}$ to be the support of data that we represent as smooth real scalar functions $f : \mathcal{M} \to \mathbb{R}$. We call these scalar functions *manifold signals*. We focus on manifold signals that have finite energy, such that $f \in L^2(\mathcal{M})$. Since $L^2(\mathcal{M})$ is a Hilbert space, it is equipped with an inner product given by

$$\langle f, g \rangle_{L^2(\mathcal{M})} = \int_{\mathcal{M}} f(x)g(x)\mathrm{d}\mu(x) \qquad (4)$$

where $\mathrm{d}\mu(x)$ is the d-dimensional volume element corresponding to measure μ. Thus, the energy of the signal f is given by $\|f\|_{L^2(\mathcal{M})}^2 = \langle f, f \rangle_{L^2(\mathcal{M})}$.

A. Laplace-Beltrami Operator

On manifolds, differentiation is implemented by the *intrinsic gradient* – a local operator acting on a neighborhood of each point $x \in \mathcal{M}$ that is homeomorphic to a d-dimensional Euclidean space. This neighborhood contains all the vectors tangent to $\mathcal{M}$ at x, which is called the tangent space of $x \in \mathcal{M}$ and is denoted as $T_x\mathcal{M}$. The disjoint union of all tangent spaces on $\mathcal{M}$ is the tangent bundle $T\mathcal{M}$. Formally, the intrinsic gradient is the operator $\nabla : L^2(\mathcal{M}) \to L^2(T\mathcal{M})$ mapping scalar functions $f(x) \in L^2(\mathcal{M})$ to *tangent vector functions* $\nabla f(x) \in L^2(T_x\mathcal{M})$. The tangent vector function $\nabla f(x)$ indicates the direction of the fastest change of signal f at point x. The adjoint of the intrinsic gradient is the *intrinsic divergence*, defined as div $: L^2(T\mathcal{M}) \to L^2(\mathcal{M})$. Interpreting the tangent vector field as the velocity field of a fluid, the intrinsic divergence can be seen as a measure of the net motion of the fluid at each point [13].

The Laplace-Beltrami (LB) operator of a manifold $\mathcal{M}$ is defined as the operator $\mathcal{L} : L^2(\mathcal{M}) \to L^2(\mathcal{M})$ given by the function composition of the intrinsic divergence and the intrinsic gradient. When considered in the local coordinates supported on $T_x\mathcal{M}$ [31], the LB operator can be written as

$$\mathcal{L}f = -\mathrm{div} \circ \nabla f = -\nabla \cdot \nabla f. \qquad (5)$$

Much like the Laplace operator in Euclidean domains (or the Laplacian matrix, in the case of graphs [32]), the LB operator measures the total variation of function f, i.e., how much the value of f at a point deviates from local average of the values of f in its surroundings [13]. Since $\mathcal{L}$, like ∇, is a local operator depending on the tangent space $T_x\mathcal{M}$ of each point $x \in \mathcal{M}$, in the following we make this dependence explicit by writing $\mathcal{L} = \mathcal{L}_x$ and $\nabla = \nabla_x$.

The LB operator plays an important role in partial differential equations (PDEs), as it governs the dynamics of the diffusion of heat over manifolds as given by the *heat equation*

$$\frac{\partial u(x,t)}{\partial t} + \mathcal{L}u(x,t) = 0. \qquad (6)$$

If $u(x,t) \in L^2(\mathcal{M})$ measures the temperature of point $x \in \mathcal{M}$ at time $t \in \mathbb{R}^+$, equation (6) can be interpreted to mean that, at point x, the rate at which the manifold "cools down" is proportional to the difference between the temperature of x and the local average of the temperature of the points in its neighborhood. With initial condition $u(x,0) = f(x)$, the solution to this equation is given by

$$u(x,t) = e^{-t\mathcal{L}}f(x), \tag{7}$$

which is the key support to implement the LB operator in the definitions proposed later.

The LB operator $\mathcal{L}$ is self-adjoint and positive-semidefinite. Considering that $\mathcal{M}$ is compact, the LB operator $\mathcal{L}$ has a real positive eigenvalue spectrum $\{\lambda_i\}_{i=1}^{\infty}$ satisfying

$$\mathcal{L}\phi_i = \lambda_i\phi_i \tag{8}$$

where ϕ_i is the eigenfunction associated with eigenvalue λ_i. The indices i are such that the eigenvalues are ordered in increasing order as $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$. In particular, for a d-dimensional manifold, we have that $\lambda_i \propto i^{2/d}$ as a consequence of Weyl's law [33]. The eigenfunctions ϕ_i are orthonormal and form a generalized eigenbasis of $L^2(\mathcal{M})$ in the intrinsic sense. Since $\mathcal{L}$ is a total variation operator, the eigenvalues λ_i can be interpreted as the canonical frequencies and the eigenfunctions ϕ_i as the canonical oscillation modes of $\mathcal{M}$. This further allows us to implement operator $\mathcal{L}$ in the spectral domain.

B. Manifold Filters

Time signals are processed by filters which compute the continuous time convolution of the input signal and the filter impulse response [17]; images and high-dimensional Euclidean signals are processed by filters implementing multidimensional convolutions [34]; and graph signals are filtered by computing graph convolutions [5]. In this paper, we define a manifold filter as the convolution of the filter impulse response $\tilde{h}(t)$ and the manifold signal f. Note that the definition of the convolution operation, denoted as $\star_{\mathcal{M}}$, leverages the heat diffusion dynamics described in (7).

Definition 1 (Manifold filter) *Let $\tilde{h}(t) : \mathbb{R}^+ \to \mathbb{R}$ and let $f \in L^2(\mathcal{M})$ be a manifold signal. The manifold filter with impulse response $\tilde{h}(t)$, denoted as $\mathbf{h}$, is given by*

$$g(x) = (\mathbf{h}f)(x) := (\tilde{h} \star_{\mathcal{M}} f)(x) := \int_0^\infty \tilde{h}(t)u(x,t)dt \tag{9}$$

where $\tilde{h} \star_{\mathcal{M}} f$ is the manifold convolution *of $\tilde{h}$ and f, and $u(x,t)$ is the solution of the heat equation* (6) *with initial condition $u(x,0) = f(x)$.*

In a slight abuse of nomenclature, in the following we will use the terms manifold filter and manifold convolution interchangeably.

From Definition 1, we see that the manifold filter operates on manifold signals $f(x)$ by (i) scaling the diffusion process (7) starting at $f(x)$ by $\tilde{h}(t)$ and (ii) aggregating the outcome of the scaled diffusion process from $t = 0$ to $t = \infty$. This definition is somewhat contrived because it does not allow computing the output of the filter directly from f. To be able to do so,

we need to substitute the solution of the heat equation (7) for $u(x,t)$ in (9). This leads to a closed-form expression of $\mathbf{h}$ that is parametric on the LB operator as shown in Proposition 1.

Proposition 1 *The manifold filter $\mathbf{h}$ is a parametric map $\mathbf{h}(\mathcal{L})$ of the LB operator $\mathcal{L}$, which is shown as*

$$g(x) = (\mathbf{h}f)(x) = \int_0^\infty \tilde{h}(t)e^{-t\mathcal{L}}f(x)dt = \mathbf{h}(\mathcal{L})f(x). \tag{10}$$

Proposition 1 tells us that manifold filters are spatial operators, since they operate directly on points $x \in \mathcal{M}$; moreover, it tells us that they are local, because they are parametrized by $\mathcal{L}$ which is itself a local operator. The exponential term $e^{-t\mathcal{L}}$ can be seen as a diffusion or shift operation akin to a time delay in a linear time-invariant (LTI) filter [17], or as the graph shift operator in a linear shift-invariant (LSI) graph filter [5]. Indeed, if we consider the manifold $\mathcal{M}$ to be the real line, the manifold filter defined in (10) recovers a LTI filter. If we consider it to be a set of points connected by a geometric graph, (10) recovers a LSI graph filter. We discuss these special cases in further detail in Appendix A and Section VI.

C. Frequency Representation of Manifold Filters

A manifold signal $f \in L^2(\mathcal{M})$ can be represented in the frequency domain of the manifold by projecting f onto the LB operator eigenbasis (8) as

$$[\hat{f}]_i = \langle f, \phi_i \rangle_{L^2(\mathcal{M})} = \int_{\mathcal{M}} f(x)\phi_i(x)d\mu(x), \tag{11}$$

where we claim that $\hat{f}$ is the *frequency representation* of the corresponding signal with $f = \sum_{i=1}^\infty [\hat{f}]_i\phi_i$.

Frequency representations are useful because they help understand the frequency behavior of the manifold filter $\mathbf{h}(\mathcal{L})$. To see this, we consider the frequency representation of the manifold filter output g in (10), which is

$$[\hat{g}]_i = \int_{\mathcal{M}} \int_0^\infty \tilde{h}(t)e^{-t\mathcal{L}}f(x)dt\phi_i(x)d\mu(x). \tag{12}$$

Rearranging the integrals and substituting $e^{-t\mathcal{L}}\phi_i = e^{-t\lambda_i}\phi_i$, we can get

$$[\hat{g}]_i = \int_0^\infty \tilde{h}(t)e^{-t\lambda_i}dt[\hat{f}]_i. \tag{13}$$

The expression relating $\hat{g}$ and $\hat{f}$ is called the *frequency response* of the filter $\mathbf{h}(\mathcal{L})$.

Definition 2 (Frequency response) *The frequency response of the filter $\mathbf{h}(\mathcal{L})$ is given by*

$$\hat{h}(\lambda) = \int_0^\infty \tilde{h}(t)e^{-t\lambda}dt. \tag{14}$$

An important consequence of Definition 2 is that, since $\hat{h}(\lambda)$ is parametric on λ, the manifold filter is pointwise in the frequency domain. This can be seen by plugging (14) into (13), and is stated explicitly in Proposition 2.

Proposition 2 *The manifold filter $\mathbf{h}(\mathcal{L})$ is pointwise in the frequency domain, which is written as*

$$[\hat{g}]_i = \hat{h}(\lambda_i)[\hat{f}]_i. \tag{15}$$

Definition 2 also emphasizes that the frequency response of a manifold filter is independent of the underlying manifold. Note that, in (14), $\hat{h}(\lambda)$ is a function of an arbitrary scalar variable λ. To obtain the frequency behavior of this filter on a given manifold $\mathcal{M}$, we need to evaluate $\hat{h}$ at the corresponding LB operator eigenvalues λ_i [cf. (8)]. If the manifold changes (or if we want to deploy the same filter on a different manifold $\mathcal{M}'$), it suffices to reevaluate $\hat{h}$ at λ_i', i.e., at the eigenvalues of the new LB operator $\mathcal{L}'$.

III. STABILITY OF MANIFOLD FILTERS WITH RESPECT TO MANIFOLD DEFORMATIONS

On the manifold $\mathcal{M}$, we define a deformation as function $\tau(x) : \mathcal{M} \to \mathcal{M}$, where $x \in \mathcal{M}$ is a point on the manifold and $\text{dist}(x, \tau(x))$ is upper bounded, i.e., $\tau(x)$ is a displaced point in the neighborhood of x. The deformation τ has a corresponding tangent map $\tau_* : T_x\mathcal{M} \to T_{\tau(x)}\mathcal{M}$ and a Jacobian matrix $J(\tau_*)$. When $\text{dist}(x, \tau(x))$ is bounded, the Frobenius norm of $J(\tau_*) - I$ can also be upper bounded, and these bounds are used to measure the size of the deformation $\tau(x)$.

Let $f : \mathcal{M} \to \mathbb{R}$ be a manifold signal. Because $\mathcal{M}$ is the codomain of $\tau(x)$, $g = f \circ \tau$ maps points $\tau(x) \in \mathcal{M}$ to $f(\tau(x)) \in \mathbb{R}$, so that the effect of a manifold deformation on the signal f is a signal perturbation leading to a new signal g supported on the same manifold. To understand the effect of this deformation on the LB operator, let $p = \mathcal{L}g$. Since p is also a signal on $\mathcal{M}$, we may define an operator $\mathcal{L}'$ mapping f directly into p,

$$p(x) = \mathcal{L}'f(x) = \mathcal{L}g(x) = \mathcal{L}f(\tau(x)). \tag{16}$$

The operator $\mathcal{L}'$ is the perturbed LB operator, which is effectively the new LB operator resulting from the deformation τ. Assuming that the gradient field is smooth, the difference between $\mathcal{L}'$ and $\mathcal{L}$ is given by the following theorem. The proof is deferred to Appendix B.

Theorem 1 *Let $\mathcal{L}$ be the LB operator of the manifold $\mathcal{M}$. Let $\tau(x) : \mathcal{M} \to \mathcal{M}$ be a manifold perturbation such that $\text{dist}(x, \tau(x)) = \epsilon$ and $J(\tau_*) = I + \Delta$ with $\|\Delta\|_F = \epsilon$. If the gradient field is smooth, it holds that*

$$\mathcal{L} - \mathcal{L}' = \mathcal{E}\mathcal{L} + \mathbf{A}, \tag{17}$$

where $\mathcal{E}$ and $\mathbf{A}$ satisfy $\|\mathcal{E}\| = O(\epsilon)$ and $\|\mathbf{A}\|_{op} = O(\epsilon)$.

Therefore, the perturbation of the LB operator incurred by a manifold deformation τ is a combination of an absolute perturbation $\mathbf{A}$ [cf. Definition 3] and a relative perturbation $\mathcal{E}\mathcal{L}$ [cf. Definition 7]. This largely simplifies our analysis of stability. Since manifold filters are parametric on $\mathcal{L}$ [cf. Proposition 2], it is sufficient to characterize their stability to deformations of the manifold by analyzing their behavior in the presence of absolute and relative LB perturbations. This is what we do in Sections III-A and III-B.

A. Stability of Manifold Filters to Absolute Perturbations

We start by analyzing the stability of manifold filters to absolute perturbations of the LB operator, which are introduced in Definition 3.

Definition 3 (Absolute perturbations) *Let $\mathcal{L}$ be the LB operator of manifold $\mathcal{M}$. An absolute perturbation of $\mathcal{L}$ is defined as*

$$\mathcal{L}' - \mathcal{L} = \mathbf{A}, \tag{18}$$

where the absolute perturbation operator $\mathbf{A}$ is self-adjoint.

Like $\mathcal{L}$, the operator $\mathcal{L}'$ resulting from the absolute perturbation of $\mathcal{L}$ is self-adjoint due to the symmetry of $\mathbf{A}$. Hence, it admits an eigendecomposition similar to (5). When the filter coefficients are fixed, the frequency response of the manifold filter (15) can be obtained by evaluating $\hat{h}(\lambda)$ at each λ_i. Thus, to understand the effect of the perturbation on the filter we need to look at how the perturbation of the LB operator changes the eigenvalues λ_i. The challenge in this case is that the spectrum of $\mathcal{L}$ is infinite-dimensional, i.e., there is an infinite (though countable) number of eigenvalue perturbations that need to be taken into account. As demonstrated by Proposition 3, however, large eigenvalues accumulate in certain parts of the real line. This suggests a strategy to partition the spectrum into finite number of partitions.

Proposition 3 *Consider a d-dimensional compact manifold $\mathcal{M} \subset \mathbb{R}^N$ and let $\mathcal{L}$ be its LB operator with eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$. Let C_1 be an arbitrary constant and C_d the volume of the d-dimensional unit ball. Let $Vol(\mathcal{M})$ denote the volume of manifold $\mathcal{M}$. For any $\alpha > 0$ and $d > 2$, there exists N_1,*

$$N_1 = \lceil (\alpha d/C_1)^{d/(2-d)} (C_d Vol(\mathcal{M}))^{2/(2-d)} \rceil \tag{19}$$

such that, for all $k > N_1$,

$$\lambda_{k+1} - \lambda_k \leq \alpha.$$

Proof. This is a direct consequence of Weyl's law [35, Chapter 1]. ∎

Given this asymptotic behavior, we can divide the eigenvalues into a finite number of partitions by placing eigenvalues that are less than $\alpha > 0$ apart from each other in groups. This spectrum separation strategy is described in Definition 4. To achieve it, we will need a specific type of manifold filter called Frequency Difference Threshold (FDT) filter as introduced in Definition 5.

Definition 4 (α-separated spectrum) *The α-separated spectrum of a LB operator $\mathcal{L}$ is defined as the partition $\Lambda_1(\alpha) \cup \ldots \cup \Lambda_N(\alpha)$ such that all $\lambda_i \in \Lambda_k(\alpha)$ and $\lambda_j \in \Lambda_l(\alpha)$, $k \neq l$, satisfy*

$$|\lambda_i - \lambda_j| > \alpha. \tag{20}$$

Definition 5 (α-FDT filter) *The α-frequency difference threshold (α-FDT) filter is defined as a filter $\mathbf{h}(\mathcal{L})$ whose frequency response satisfies*

$$|\hat{h}(\lambda_i) - \hat{h}(\lambda_j)| \leq \delta_k \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\alpha) \tag{21}$$

with $\delta_k \leq \delta$ for $k = 1, \ldots, N$.

In the α-separated spectrum, eigenvalues $\lambda_i \in \Lambda_k(\alpha)$ and $\lambda_j \in \Lambda_l(\alpha)$ in different sets ($k \neq l$) are at least α away from each other. Conversely, eigenvalues $\lambda_i, \lambda_j \in \Lambda_k(\alpha)$ are no more than α apart. This partitioning creates several eigenvalue groups
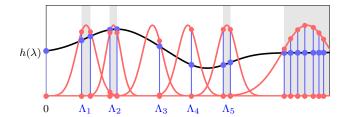
Figure 1: Illustration of an α-FDT filter. The x-axis stands for the spectrum with each sample representing an eigenvalue. The gray shaded areas show the grouping of the eigenvalues according to Definition 4. The red lines show a set of α-FDT filters that can discriminate each eigenvalue group.

spaced by at least α. Note that the sets $\Lambda_k(\alpha)$ can have any size and, in particular, they can be singletons.

The partitioning of the spectrum described in Definition 4 is achieved by an α-FDT filter. This filter separates the spectrum of the manifold by assigning similar frequency responses— that deviate no more than δ_k from each other—to eigenvalues $\lambda_i \in \Lambda_k(\alpha)$, $1 \leq k \leq N$. In other words, the α-FDT filter does not discriminate between eigenvalues $\lambda_i, \lambda_j \in \Lambda_k(\alpha)$. Importantly, the δ_k in Definition 4 are finite, so that they can be bounded by some δ.

To obtain manifold filters that are stable to absolute perturbations of $\mathcal{L}$, we also need these filters to be Lipschitz continuous as shown in Definition 6.

Definition 6 (Lipschitz filter) *A filter is A_h-Lispchitz if its frequency response is Lipschitz continuous with Lipschitz constant A_h, i.e,*

$$|\hat{h}(a) - \hat{h}(b)| \leq A_h|a - b| \text{ for all } a, b \in (0, \infty). \quad (22)$$

Between the eigenvalue groups, the filters that we consider are assumed to be A_h-Lipschitz continuous. This means that, in regions of the spectrum where the $\Lambda_k(\alpha)$ are singletons, the filter can vary with slope at most A_h as shown in Figure 1. Note that we can always construct convolutional filters (14) that are both Lipschitz continuous and α-FDT.

Under mild assumptions on the amplitude of the frequency response $\hat{h}$ (Assumption 1), it can be shown that Lipschitz continuous α-FDT filters are stable to absolute perturbations of the LB operator. This result is stated in Theorem 2.

Assumption 1 (Non-amplifying filters) *The filter frequency response $\hat{h} : \mathbb{R}^+ \to \mathbb{R}$ is non-amplifying. I.e., for all $\lambda \in (0, \infty)$, $\hat{h}$ satisfies $|\hat{h}(\lambda)| \leq 1$.*

Note that this assumption is rather reasonable, because the filter frequency response $\hat{h}(\lambda)$ can always be normalized.

Theorem 2 (Manifold filter stability to absolute perturbations) *Consider a manifold $\mathcal{M}$ with LB operator $\mathcal{L}$. Let $\mathbf{h}(\mathcal{L})$ be a α-FDT manifold filter [cf. Definition 5] and A_h-Lipschitz [cf. Definition 6]. Consider an absolute perturbation $\mathcal{L}' = \mathcal{L} + \mathbf{A}$*

of the LB operator $\mathcal{L}$ [cf. Definition 3] where $\|\mathbf{A}\| = \epsilon < \alpha$. Then, under Assumption 1 it holds that

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\|_{L^2(\mathcal{M})} \leq$$
$$\left(\frac{\pi N_s \epsilon}{\alpha - \epsilon} + A_h \epsilon + 2(N - N_s)\delta\right) \|f\|_{L^2(\mathcal{M})}, \quad (23)$$

where N is the size of the α-separated spectrum partition [cf. Definition 4] and N_s is the number of singletons.

Proof. See Appendix C. ∎

Provided that $\epsilon \ll \alpha$, FDT filters are thus stable to absolute perturbations of the LB operator $\mathcal{L}$. The stability bound depends on (i) the continuity of the FDT filter as measured by the Lipschitz constant A_h and (ii) its frequency difference threshold α, which affects the bound directly as well as indirectly through the number of partitions N. Note that this bound consists of three terms. The first corresponds to the difference between the eigenfunctions of $\mathcal{L}$ and $\mathcal{L}'$, which affects the stability bound by changing projection directions. The second stems from the distance between the original and perturbed eigenvalues. Finally, the third reflects the bounded fluctuation of the filter frequency response within the same eigenvalue group.

The bound in Theorem 2 can be simplified by setting $\delta = \pi\epsilon/(2\alpha - 2\epsilon)$ as in Corollary 1.

Corollary 1 *Setting $\delta = \pi\epsilon/(2\alpha - 2\epsilon)$, under the same assumptions of Theorem 2 it holds that*

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\|_{L^2(\mathcal{M})} \leq \left(\frac{\pi N}{\alpha - \epsilon} + A_h\right)\epsilon\|f\|_{L^2(\mathcal{M})}. \quad (24)$$

A particular case of Theorem 2, the simplified stability bound in Corollary 1 is helpful to understand the effect of the filter spectrum on stability as well as of the size of the perturbation. In particular, from Corollary 1 we can tell that the filter is more stable if the Lipschitz constant A_h is small and the frequency difference threshold α is large. On the other hand, small A_h and large α mean that the filter is less discriminative. This reveals a stability-discriminability trade-off where discriminability should be understood as the ability to tell frequencies apart. In other words, we propose α-FDT filter to maintain the stability by trying to discriminate only eigenvalue groups instead of every single eigenvalue. Importantly, this trade-off is not related to the magnitude of the frequencies that the filters amplify (as is the case in, e.g., [19], [36]).We will keep observing this trade-off throughout the stability analysis of both manifold filters and MNNs. More details about this characteristic will be discussed in Section V.

B. Stability of Manifold Filters to Relative Perturbations

Relative perturbations of the LB operator are defined similarly as follows.

Definition 7 (Relative perturbations) *Let $\mathcal{L}$ be the LB operator of manifold $\mathcal{M}$. A relative perturbation of $\mathcal{L}$ is defined as*

$$\mathcal{L}' - \mathcal{L} = \mathcal{E}\mathcal{L}, \quad (25)$$

where the relative perturbation term $\mathcal{E}\mathcal{L}$ is self-adjoint.

Like absolute perturbations, relative perturbations also perturb the eigenvalues and eigenfunctions of $\mathcal{L}$. However, in the case of relative perturbations, the perturbations to the eigenvalues are proportional to their absolute values [cf. Lemma 3 in the supplementary material]. Relative perturbations thus require a different spectrum separation strategy to guarantee stability. With relative perturbations, larger eigenvalues are impacted with larger perturbation values, which can send eigenvalues originally α-close to each other to different groups as Figure 2. Therefore, we will need a different type of filter implementing a different type of spectrum separation. Our strategy is inspired by Proposition 4, which is another variation of Weyl's law.

Proposition 4 *Let $\mathcal{M}$ be a d-dimensional compact embedded manifold in $\mathbb{R}^N$ with LB operator $\mathcal{L}$, and let $\{\lambda_k\}_{k=1}^\infty$ denote the eigenvalues of $\mathcal{L}$. Let C_1 denote an arbitrary constant. For any $\gamma > 0$, there exists N_2 given by*

$$N_2 = \lceil (C_1(\gamma + 1)^{d/2} - 1)^{-1} \rceil \tag{26}$$

such that, for all $k > N_2$, it holds that

$$\lambda_{k+1} - \lambda_k \leq \gamma \lambda_k.$$

Proof. This is a direct consequence of Weyl's law [35, Chapter 1]. ∎

Hence, to enforce stability we need to separate the spectrum relatively to the ratio between neighboring eigenvalues. This partitioning is called the γ-separated spectrum and formalized in Definition 8. The γ-separated spectrum is achieved by the so-called Frequency Ratio Threshold (FRT) filters. We introduce them in Definition 9.

Definition 8 (γ-separated spectrum.) *The γ-separated spectrum of a LB operator $\mathcal{L}$ is defined as the partition $\Lambda_1(\gamma) \cup \ldots \cup \Lambda_M(\gamma)$ such that all $\lambda_i \in \Lambda_k(\gamma)$ and $\lambda_j \in \Lambda_l(\gamma)$, $k \neq l$, satisfy*

$$\left| \frac{\lambda_i}{\lambda_j} - 1 \right| > \gamma. \tag{27}$$

Definition 9 (γ-FRT filter.) *The γ-frequency ratio threshold (γ-FRT) filter is defined as a filter $\mathbf{h}(\mathcal{L})$ whose frequency response satisfies*

$$|\hat{h}(\lambda_i) - \hat{h}(\lambda_j)| \leq \delta_k, \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\gamma) \tag{28}$$

with $\delta_k \leq \delta$ for $k = 1, 2 \ldots, M$.

In the γ-separated spectrum, the sets $\Lambda_k(\gamma)$ are built based on eigenvalue distances *relative* to the eigenvalues' magnitudes and weighted by the parameter γ. Eigenvalues $\lambda_j \in \Lambda_k(\gamma)$ and $\lambda_i \in \Lambda_l(\gamma)$ in different groups (i.e., $k \neq l$) are at least $\gamma \min(\lambda_i, \lambda_j)$ apart from each other. This means that, for $\lambda_i, \lambda_{i+1} \in \Lambda_k(\gamma)$, $\lambda_{i+1} - \lambda_i \leq \gamma \lambda_i$.

The γ-FRT filter achieves the spectrum separation in Definition 8 by giving eigenvalues $\lambda_i, \lambda_j \in \Lambda_k(\gamma)$ very similar frequency responses differing by at most plus or minus $\delta_k < \delta$. Meanwhile, eigenvalues belonging to different sets $\Lambda_k(\gamma)$ and $\Lambda_l(\gamma)$, $k \neq l$, are treated independently, and their frequency response can vary a lot.

To make a manifold filter stable to relative perturbations of the LB operator, we need a further restriction on their continuity.
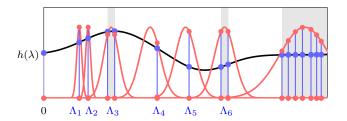

Figure 2: Illustration of a γ-FRT filter. The x-axis stands for the spectrum with each sample representing an eigenvalue. The gray shaded area shows the grouping of the eigenvalues according to Definition 8. The red lines show a set of α-FDT filters that can discriminate each eigenvalue group.

Lipschitz continuity [cf. Definition 6] is not enough because in a Lipschitz filter the difference in frequency response for a perturbed eigenvalue grows with the eigenvalue magnitude, since the eigenvalue perturbation is relative. Therefore, we need our filters to be *integral Lipschitz* continuous as is described in Definition 10.

Definition 10 (Integral Lipschitz filter) *A filter is integral Lipschitz continuous with constant B_h if its frequency response is given by*

$$|\hat{h}(a) - \hat{h}(b)| \leq \frac{B_h|a - b|}{(a + b)/2} \text{ for all } a, b \in (0, \infty). \tag{29}$$

Integral Lipschitz filters can be seen as Lipschitz filters with variable Lipschitz constant, which decreases with λ. E.g., on the interval (a, b), the filter in Definition 10 behaves as a Lipschitz filter with Lipschitz constant $2B_h/(a + b)$. When a and b are close, this condition can be approximated by $|a\hat{h}'(a)| \leq B_h$. This implies that the filter function flattens for high-frequency eigenvalues as shown in Figure 2.

Under Assumption 1, integral Lipschitz γ-FRT filters are stable to relative perturbations as stated in Theorem 3.

Theorem 3 (Manifold filter stability to relative perturbations) *Consider a manifold $\mathcal{M}$ with LB operator $\mathcal{L}$. Let $\mathbf{h}(\mathcal{L})$ be a γ-FRT filter with $\delta = \pi\epsilon/(2\gamma - 2\epsilon + 2\gamma\epsilon)$ [cf. Definition 9] and B_h-integral Lipschitz [cf. Definition 10]. Consider a relative perturbation $\mathcal{L}' = \mathcal{L} + \mathbf{E}\mathcal{L}$ of the LB operator $\mathcal{L}$ [cf. Definition 7] where $\|\mathbf{E}\| = \epsilon < \gamma$. Then, under Assumption 1 it holds that*

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\|_{L^2(\mathcal{M})} \leq \left(\frac{\pi M \epsilon}{\gamma - \epsilon + \gamma\epsilon} + \frac{2B_h\epsilon}{2 - \epsilon} \right) \|f\|_{L^2(\mathcal{M})} \tag{30}$$

where M is the size of the γ-separated spectrum partition [cf. Definition 8].

Proof. See Section B in supplementary material. ∎

When ϵ is sufficiently small ($\epsilon \ll \min(\gamma, 2)$, which is typically the case with deformations such as the one in Theorem 1), the denominators on the right hand side of (30) are approximately equal to γ and 2 respectively. Hence, γ-FRT integral Lipschitz filters are stable to relative perturbations of the LB operator. Besides appearing in the bound in Theorem 3), the frequency ratio threshold γ also affects stability indirectly through the partition size M. With a larger γ, fewer eigenvalues

will be in singleton sets, thus decreasing M and improving stability. A smaller integral Lipschitz constant B_h also increases stability. However, small B_h and large γ make for smoother filters which in turn lead to a less discriminative manifold filter. Therefore, integral Lipschitz γ-FRT filters also exhibit a trade-off between discriminality and stability.

Remark 1 By comparing the illustrations of α-FDT filter (Definition 5) and γ-FRT filter (Definition 9) in Figure 1 and Figure 2, we see that in practice these filters have a similar frequency behavior because, due to Weyl's law [cf. Proposition 3 and 4], high frequency components will eventually be grouped in the same group and thus share similar frequency responses. Therefore, the main difference between these filters is their effects on the low-frequency components. In the low frequency spectrum, the eigengaps $\lambda_{i+1} - \lambda_i$ may be smaller than the difference threshold α, but larger than the relative ratio threshold $\gamma \lambda_i$ due to λ_i being small. However, for appropriate values of γ a filter may be both FDT and FRT. This will be shown in Section IV.

IV. STABILITY OF MANIFOLD NEURAL NETWORKS

Manifold neural networks (MNNs) are deep convolutional architectures comprised of L layers, where each layer consists of two components: a convolutional filterbank and a pointwise nonlinearity. At each layer $l = 1, 2, \ldots, L$, the convolutional filters map the incoming F_{l-1} features from layer $l-1$ into F_l intermediate linear features given by

$$y_l^p(x) = \sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q(x), \qquad (31)$$

where $\mathbf{h}_l^{pq}(\mathcal{L})$ is the filter mapping the q-th feature from layer $l-1$ to the p-th feature of layer l as in (15), for $1 \le q \le F_{l-1}$ and $1 \le p \le F_l$. The intermediate features are then processed by a pointwise nonlinearity $\sigma : \mathbb{R} \to \mathbb{R}$ as

$$f_l^p(x) = \sigma\left(y_l^p(x)\right). \qquad (32)$$

The nonlinearity σ processes each feature individually and we further make an assumption on its continuity as follows.

Assumption 2 (Normalized Lipschitz activation functions)
The activation function σ is normalized Lipschitz continuous, i.e., $|\sigma(a) - \sigma(b)| \le |a - b|$, with $\sigma(0) = 0$.

Note that this assumption is rather reasonable, since most common activation functions (e.g., the ReLU, the modulus and the sigmoid) are normalized Lipschitz by design.

At the first layer of the MNN, the input features are the input data f^q for $1 \le q \le F_0$. At the output of the MNN, the output features are given by the outputs of the L-th layer, i.e., f_L^p for $1 \le p \le F_L$. To represent the MNN more succinctly, we may gather the impulse responses of the manifold convolutional filters $\mathbf{h}_l^{pq}$ across all layers in a function set $\mathbf{H}$, and define the MNN map $\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$. This map emphasizes that the MNN is parameterized by both the filter functions and the LB operator $\mathcal{L}$. We next will analyze the stability of $\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$ with respect to perturbations on the underlying manifold.

A. Stability of MNNs to LB Operator Perturbations

MNNs inherit stability to perturbations of the LB operator from the manifold filters that compose the filterbanks in each one of their layers. This result is stated in general form—encompassing both absolute and relative perturbations—in the following theorem.

Theorem 4 (MNN stability) *Consider a compact embedded manifold $\mathcal{M}$ with LB operator $\mathcal{L}$. Let $\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$ be an L-layer MNN on $\mathcal{M}$ (32) with $F_0 = F_L = 1$ input and output features and $F_l = F, l = 1, 2, \ldots, L-1$ features per layer. The filters $\mathbf{h}(\mathcal{L})$ and nonlinearity functions satisfy Assumptions 1 and 2 respectively. Let $\mathcal{L}'$ be the perturbed LB operator [cf. Definition 3 or Definition 7] with $\max\{\alpha, 2, |\gamma/1-\gamma|\} \gg \epsilon$. If the manifold filters satisfy $\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\|_{L^2(\mathcal{M})} \le C_{per}\epsilon\|f\|_{L^2(\mathcal{M})}$, it holds that*

$$\|\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f) - \mathbf{\Phi}(\mathbf{H}, \mathcal{L}', f)\|_{L^2(\mathcal{M})} \le LF^{L-1}C_{per}\epsilon\|f\|_{L^2(\mathcal{M})}.$$

Proof. See Section B in supplementary material. ∎

Theorem 4 reflects that the stability of the MNN is affected by the hyperparameters of the MNN architecture and the stability constant of the manifold filters C_{per}. More explicitly, the stability bound grows linearly with the number of layers L and exponentially with the number of features F where the rate is determined by L. This stability result also shows that there is a linear dependence on the stability constant C_{per} of manifold filters $\mathbf{h}(\mathcal{L})$ and the perturbation size ϵ. As we have shown in Section III-A and III-B, the stability constant is determined by the form of the perturbations (Definition 3 or Definition 7) as well as the spectrum separation achieved by the specific manifold filters (Definition 5 or Definition 9) with corresponding Lipschitz conditions (Definition 6 or Definition 10). We address the specific cases as follows.

Proposition 5 *With the same conditions as Theorem 4, consider the following perturbation models.*

1) *If the perturbed LB operator $\mathcal{L}'$ is an absolute perturbation, i.e., $\mathcal{L}' = \mathcal{L} + \mathbf{A}$ [cf. Definition 3] with $\|\mathbf{A}\| = \epsilon$ and the manifold filters $\mathbf{h}(\mathcal{L})$ are α-FDT [cf. Definition 5] with $\alpha \gg \epsilon$ and A_h-Lipschitz continuous [Definition 6] with $\delta = \pi\epsilon/(2\alpha)$, we have*

$$C_{per} = \frac{\pi N}{\alpha} + A_h, \qquad (33)$$

where N is the size of the α-separated spectrum partition [cf. Definition 4].

2) *If the perturbed LB operator $\mathcal{L}'$ is a relative perturbation, i.e. $\mathcal{L}' = \mathcal{L} + \mathbf{E}\mathcal{L}$ [cf. Definition 7] with $\|\mathbf{E}\| = \epsilon$, and the manifold filters $\mathbf{h}(\mathcal{L})$ are γ-FRT [cf. Definition 9] with $\gamma/(1 - \gamma) \gg \epsilon$ and B_h-integral Lipschitz continuous [Definition 10] with $\delta = \pi\epsilon/(2\gamma)$, we have*

$$C_{per} = \frac{\pi M}{\gamma} + B_h, \qquad (34)$$

where M is the size of the γ-separated spectrum partition [cf. Definition 8].

Proof. The conclusions follow directly from Theorem 4 combined with Theorem 2 or Theorem 3 under the corresponding assumptions. ∎

Combining Theorem 4 with Proposition 5, we observe that α-FDT manifold filters with Lipschitz continuity can be composed to construct MNNs which are stable to absolute perturbations; while γ-FRT manifold filters with integral Lipschitz continuity can be composed to construct MNNs which are stable relative perturbations of the LB operator. Explicitly, by inserting the stability constant C_{per} in (23), we see that other than the perturbation size ϵ, there are three terms that determine the stability of MNNs. The first term is LF^{L-1}, which, as we have already discussed, is decided by the number of layers and filters in the MNN architecture. This term arises due to the propagation of the underlying operator perturbations across all the manifold filters in all layers of the MNN. The second term is $\pi N/\alpha$ or $\pi M/\gamma$, which results from the deviations of the eigenfunctions as well as from the frequency response variations within the same eigenvalue partition. Finally, the third term, A_h or B_h, is given by the Lipschitz or integral Lipschitz constants which are decided during the filter design or the training process. It is important to note that the stability constant C_{per} brings along the trade-off between stability and discriminability. However, unlike manifold filters, MNNs can be both stable and discriminative. This arises from the effects of nonlinear activation functions, as we discuss in further detail in Section V.

B. Stability of MNNs to Manifold Deformations

In Theorem 4 and Proposition 5, we established the conditions under which MNNs are stable to either absolute or relative perturbations of the LB operator as defined in Definitions 3 and 7. Since a manifold deformation $\tau(x) : \mathcal{M} \to \mathcal{M}$, with $\text{dist}(x, \tau(x)) = \epsilon$ and $J(\tau_*) = I + \Delta$, translates into both an absolute and a relative perturbation of the Laplace-Beltrami operator, in order for the MNN to be stable to this deformation we need to meet all of these conditions in items 1 and 2 of Proposition 5. I.e., the manifold filters need to be both α-FDT and γ-FRT, and both Lipschitz continuous and integral Lipschitz continuous. The spectrum can be made to be both α-separated and γ-separated by making sure the eigenvalues in different partitions satisfy both (20) and (27). Assuming that all of these conditions are met, we can state our main result—that MNNs are stable to deformations of the manifold—as follows.

Theorem 5 *Let $\mathcal{M}$ be a compact embedded manifold with LB operator $\mathcal{L}$ and f be manifold signal. We construct $\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$ as a MNN on $\mathcal{M}$ (32) where the filters $\mathbf{h}(\mathcal{L})$ are α-FDT [cf. Definition 5], α/λ_1-FRT [cf. Definition 9], A_h-Lipschitz [cf. Definition 6] and B_h-integral Lipschitz [cf. Definition 10]. Consider a deformation on $\mathcal{M}$ as $\tau(x) : \mathcal{M} \to \mathcal{M}$ where $\text{dist}(x, \tau(x)) = \epsilon$ and $J(\tau_*) = I + \Delta$ with $\|\Delta\|_F = \epsilon$ and $\epsilon \ll \min(\alpha/\lambda_1, \alpha, 2)$. Then under Assumptions 1 and 2 it holds that*

$$\|\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f) - \mathbf{\Phi}(\mathbf{H}, \mathcal{L}', f)\|_{L^2(\mathcal{M})} = O(\epsilon)\|f\|_{L^2(\mathcal{M})}. \quad (35)$$

Together, Theorem 1 and Theorem 4 imply that MNNs are stable to the manifold deformations υ introduced in the beginning of this section. This is because these deformations spawn a perturbation of the LB operator that consists of both an absolute and a relative perturbation. For stability to hold, the filters that make up the layers of the MNN need to be

α-FDT [cf. Definition 5], γ-FRT [cf. Definition 9], Lipschitz [cf. Definition 6] and integral Lipschitz [cf. Definition 10]. We can propose an easier special case to relate α and γ by utilizing the spectrum property of LB operator. By setting the α-FDT filter with $\alpha = \gamma\lambda_1$, eigenvalues $\lambda_i, \lambda_{i+1} \in \Lambda_k(\alpha)$ would lead to $\lambda_i, \lambda_{i+1} \in \Lambda_l(\gamma)$ due to the fact that

$$\lambda_{i+1} - \lambda_i \leq \alpha = \gamma\lambda_1 \leq \gamma\lambda_i, \quad (36)$$

with λ_1 indexed as the smallest eigenvalue in the spectrum. The requirement that the filter be α-FDT can be removed as long as $\lambda_1 > 0$ and $\alpha = \gamma\lambda_1$, since a γ-FRT filter is always $\gamma\lambda_1$-FDT, i.e. α-FDT.

V. DISCUSSIONS

Stability vs. discriminability tradeoff. In both stability theorems for manifold filters (Theorems 2, 3) and in the stability theorem for MNNs (Theorem 4), the stability bounds depend on the frequency partition threshold (α or γ), the number of total partitions (N or M) and the Lipschitz continuity constant (A_h or B_h). The frequency partition threshold and the number of partitions have a combined effect on stability. As indicated by Definitions 4 and 8, a larger frequency threshold leads to a smaller number of singletons, as eigenvalues that would otherwise be separated for small thresholds end up being grouped when the threshold is large. While a large frequency threshold results in a larger number of partitions that contain more than one eigenvalue, the total number of partitions (N or M) either stays the same or decreases because the number of eigenvalues does not exceed the number of partitions [cf. Proposition 3 or 4]. Thus, a larger frequency threshold and a smaller number of partitions both lead to a smaller stability bound. Simultaneously, a large frequency threshold makes the spectrum separated more sparsely. Therefore, a large number of eigenvalues are amplified in a similar manner, which makes the filter function less discriminative. The Lipschitz constant (A_h or B_h) affects stability and discriminability in similar ways. Smaller Lipschitz constants decrease the stability bound, but lead to smoother filter functions which give similar frequency responses to different eigenvalues. Hence, in both manifold filters and MNNs we observe a trade-off between stability and discriminability. Nevertheless, in MNNs this trade-off is alleviated due to the presence of nonlinearities as discussed below.

Pointwise nonlinearity. As demonstrated by Propositions 3 and 4, large eigenvalues of LB operator tend to be grouped together in one large group and share similar frequency responses. This is part of the reason why manifold filters have a stability-discriminability tradeoff, which implies that they cannot be stable and discriminative at the same time. However, in MNNs this problem is circumvented with the addition of nonlinearities. Nonlinearities have the effect of scattering the spectral components all over the eigenvalue spectrum. In the MNN, they mix the frequency components by spilling spectral components associated with the large eigenvalues onto the smaller eigenvalues, where they can then be discriminated by the manifold filters in the following layer. This is consistent with the role of nonlinear activation functions in graph neural

networks (GNNs) n [19], which can be see as instantiations of MNNs on discrete samples of the manifold as further discussed in Section VI.

Comparison with graphons. The graphon is another infinite-dimensional model that can represent the limit of convergent sequences of graphs, and a series of works have proved stability of graphon neural networks and the transferability of GNNs sampled from them [23], [25]–[27], [36]. Manifolds are however more powerful, because they can represent the limit of graphs with both bounded and unbounded degrees [37]—the graphon is only the limit of sequences of dense graphs [38]. Moreover, embedded manifolds in high-dimensional spaces are more realistic geometric models in a number of application scenarios, such as point clouds, 3D shape segmentation and classification. Other important differences are that (i) the stability analysis on graphon models in [24], [36] focuses on deformations to the adjacency matrix of the graph, which can be translated directly as perturbations of the graphon operator, and that (ii) in the case of graphons, only an absolute perturbation model makes sense since given that the graphon spectrum is bounded a relative perturbation can always be bounded by an absolute perturbation. Meanwhile, deformations to the manifold domain translate into a combination of absolute and relative perturbations of the LB operator, and the fact that the LB operator spectrum is unbounded makes the effects of absolute and relative perturbations distinct, especially in the high-frequency domain.

VI. FROM MANIFOLD NEURAL NETWORKS TO GRAPH NEURAL NETWORKS

MNNs are built from manifold convolutional filters (Definition 1) operating on a continuous manifold and over an infinite time horizon. This makes it impractical to implement directly the architecture described by (32) in applications. In this section, we discuss how MNNs are implemented in practice over a set of discrete samples from the manifold in a finite and discrete time frame.

A. Discretization in the Space Domain

In practice, the explicit form of the manifold and of its LB operator are unknown. What we typically have access to is a point cloud representation of the manifold, i.e., a discrete set of sampling points. From these points' coordinates, the structure of the manifold is approximated by a geometric or a nearest neighbor graph [28], [37], [39]. The LB operator is then approximated by the graph Laplacian, which can be shown to converge to the LB operator as the number of sampling points grows [39] [28].

Explicitly, suppose that $X = \{x_1, x_2, \ldots, x_n\}$ is a set of n points sampled i.i.d. from measure μ of manifold $\mathcal{M}$, which is embedded in $\mathbb{R}^N$. We can construct a complete weighted symmetric graph $\mathbf{G}_n$ by taking the sampled points to be the vertices of the graph and setting the edge weights based on the Euclidean distance between pairs of points. Specifically, the weight w_{ij} associated with edge (i, j) is given by

$$w_{ij} = \frac{1}{n} \frac{1}{t_n (4\pi t_n)^{k/2}} \exp\left(-\frac{\|x_i - x_j\|^2}{4t_n}\right), \quad (37)$$

where $\|x_i - x_j\|$ is the Euclidean distance between points x_i and x_j while t_n is a parameter associated with the chosen Gaussian kernel [37]. The adjacency matrix $\mathbf{A}_n \in \mathbb{R}^{n \times n}$ is thus defined as $[\mathbf{A}_n]_{ij} = w_{ij}$ for $1 \leq i, j \leq n$ and the corresponding graph Laplacian matrix $\mathbf{L}_n$ [40] is given by

$$\mathbf{L}_n = \text{diag}(\mathbf{A}_n \mathbf{1}) - \mathbf{A}_n. \quad (38)$$

We interpret $\mathbf{L}_n$ the Laplacian operator of the constructed graph $\mathbf{G}_n$. Similarly, we define a uniform sampling operator $\mathbf{P}_n : L^2(\mathcal{M}) \to L^2(\mathbf{G}_n)$ to sample manifold signals. Given a manifold signal f, we can use operator $\mathbf{P}_n$ to sample graph signals $\mathbf{x}_n \in \mathbb{R}^n$ as

$$\mathbf{x}_n = \mathbf{P}_n f \text{ with } [\mathbf{x}_n]_i = f(x_i), \quad x_i \in X, \quad (39)$$

where the i-th entry of the graph signal $\mathbf{x}_n$ is the manifold signal f evaluated at the sample point x_i.

In Proposition 1, we have shown that the manifold filter $\mathbf{h}$ is parametric on the LB operator. Therefore, we can also parameterize $\mathbf{h}$ by the discrete graph Laplacian operator $\mathbf{L}_n$, i.e.,

$$\mathbf{z}_n = \int_0^\infty \tilde{h}(t) e^{-t\mathbf{L}_n} dt \mathbf{x}_n = \mathbf{h}(\mathbf{L}_n)\mathbf{x}_n, \ \mathbf{x}_n, \mathbf{z}_n \in \mathbb{R}^n, \quad (40)$$

where $\mathbf{z}_n$, the output of the filter, is now a discrete graph signal. By cascading these discrete filters operated on graph $\mathbf{G}_n$ and pointwise nonlinearities layer after layer, we can then approximate the MNN on $\mathbf{G}_n$ as

$$\mathbf{x}_l^p = \sigma\left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q\right), \quad (41)$$

where $\mathbf{h}_l^{pq}(\mathbf{L}_n)$ maps the q-th feature in the $l-1$-th layer to the p-th feature in the l-th layer, $1 \leq q \leq F_{l-1}$ and $1 \leq p \leq F_l$, and F_l denotes the number of features in the l-th layer (we have dropped the subscript n in $\mathbf{x}_l^p$ and $\mathbf{x}_{l-1}^q$ for simplicity). After gathering the filter functions in the set $\mathbf{H}$, this neural network can be represented more succinctly as $\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{x})$.

Equation (41) is a consistent approximation of the MNN because, as n goes to infinity, the discrete graph Laplacian operator $\mathbf{L}_n$ of the graph $\mathbf{G}_n$ converges to the LB operator $\mathcal{L}$ of the manifold $\mathcal{M}$, and the sampled graph signal $\mathbf{x}_n$ converges to the manifold signal f [37]. These facts combinely imply that the output of the neural network on the graph $\mathbf{G}_n$ converges to the output of the neural network on the continuous manifold as stated in the following.

Proposition 6 *Let $X = \{x_1, x_2, \ldots x_n\}$ be n points sampled i.i.d. from measure μ of d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$, with corresponding sampling operator $\mathbf{P}_n$ (39). Let $\mathbf{G}_n$ be a discrete graph approximation of $\mathcal{M}$ constructed from X as in (37) with $t_n = n^{-1/(d+2+\alpha)}$ and $\alpha > 0$. Let $\mathbf{\Phi}(\mathbf{H}, \cdot, \cdot)$ be a neural network parametrized either by the LB operator $\mathcal{L}$ of the manifold $\mathcal{M}$ (32) or the graph Laplacian operator $\mathbf{L}_n$ of $\mathbf{G}_n$. It holds that*

$$\lim_{n \to \infty} \|\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_n f) - \mathbf{P}_n \mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)\|_{L^2(\mathbf{G}_n)} = 0, \quad (42)$$

with the limit taken in probability.

Proof. See Section D in supplementary material. ∎

This proposition provides theoretical support to state that neural networks constructed from the discrete Laplacian $\mathbf{L}_n$ converge to MNN and thus can inherit the stability properties of the MNN.

B. Discretization in the Time Domain

In order to learn an MNN (32), we need to learn the manifold convolutional filters $\mathbf{h}_l^{pq}$. This means that we need to learn the impulse responses $\tilde{h}(t)$ in Definition 1. However, learning continuous functions $\tilde{h}$ is computationally infeasible, so we sample $\tilde{h}$ over fixed intervals of duration T_s and parameterize the filter with coefficients $h_k = \tilde{h}(kT_s)$, $k = 0, 1, 2 \ldots$. Setting the sampling interval to $T_s = 1$ for simplicity, the discrete-time manifold convolution can be written as

$$\mathbf{h}(\mathcal{L})f(x) = \sum_{k=0}^{\infty} h_k e^{-k\mathcal{L}} f(x) \tag{43}$$

where $\{h_k\}_{k=0}^{\infty}$ are called the filter coefficients or taps.

Yet, learning (44) is still impractical because there is an uncountable number of parameters h_k. To address this, we fix a time horizon of K time steps and rewrite (44) as

$$\mathbf{h}(\mathcal{L})f(x) = \sum_{k=0}^{K-1} h_k e^{-k\mathcal{L}} f(x) \tag{44}$$

which can be seen as a finite impulse response (FIR) filter with shift operator $e^{-\mathcal{L}}$. Indeed, the frequency response of this filter [cf. Proposition 2] is given by

$$\hat{h}(\lambda) = \sum_{k=0}^{K-1} h_k e^{-k\lambda}. \tag{45}$$

Combining (40) and (44), we can bring the discretization over the spatial and time domains together to rewrite the convolution operation on the discretized manifold and in the discrete-time domain, explicitly,

$$\mathbf{z}_n = \mathbf{h}(\mathbf{L}_n)\mathbf{x}_n = \sum_{k=0}^{K-1} h_k e^{-k\mathbf{L}_n} \mathbf{x}_n. \tag{46}$$

Equation (46) recovers the definition of the graph convolution [5] with graph shift operator $e^{-\mathbf{L}_n}$. This means that in practice we implement MNNs as graph neural networks (GNNs). Therefore, the stability behavior of the GNN can be seen as a proxy for the stability behavior of the MNN. We will leverage this idea in the numerical experiments of Section VII.

VII. NUMERICAL EXPERIMENTS

Dataset. We evaluate our MNN stability results on the Model-Net10 [41] classification problem. The dataset contains 3991 meshed CAD models from 10 categories for training and 908 models for testing. For each model, 300 points are uniformly randomly sampled from all points of the model to form the point cloud. Each point is characterized by the 3D coordinates as features. We formulate the problem by modeling a dense graph neural network model to approximate MNN. Each node in the graph can be modeled as the sampling point and each edge weight is constructed based on the distance between each pair of nodes. In this work our goal is to identify the CAD model for chairs as is illustrated in Figure 3 with the models for chair labeled as 1 and the others as 0. We deform the underlying manifold structure by adding random perturbations to the coordinates of the sampling points. By comparing the differences of the classification error rates, we aim to show that MNNs with Lipschitz continuous and integral Lipschitz continuous manifold filters are stable via looking into the performance of the approximated GNNs.

Neural network architectures. We build dense graphs to approximate the point cloud models. We use the coordinates of each point as node features. By connecting a point with all the other points in the point cloud, the edge weight is defined based the distance between every two points and a Gaussian kernel. The Laplacian matrix is calculated for each input point cloud model. We implement different architectures, including Graph Filters (GF) and Graph Neural Networks (GNN) with 1 and 2 layers, to solve the classification problem. The architectures with a single layer contain $F_0 = 3$ input features which are the 3d coordinates of each point, $F_1 = 64$ output features and $K = 5$ filter taps. While the architectures with 2 layers has another layer with $F_2 = 32$ features and 5 filter taps. We use the ReLU as nonlinearity. The learned graph filters are not regularized in architectures with 'NoPel' while graph filters in the other architectures are both Lipschitz and integral Lipschitz. All architectures also include a linear readout layer mapping the final output features to a binary scalar that estimates the classification.

Discriminability experiment. We train all the architectures with an ADAM optimizer [42] with learning rate set as 0.005 and decaying factors as 0.9, 0.999 by minimizing the entropy loss. The training point cloud models are divided in batches of 10 over 40 epochs. We run 5 random point samplings for all the architectures and we show the average classification error rates across these realizations as well as the standard deviation in Table I. We can observe that with the use of non-linearity, Graph Neural Networks perform better compared with Graph Filters. Architectures with more layers learn more accurate models which also leads to better performances.

Architecture	error rates
GNN1Ly	$8.04\% \pm 0.88\%$
GNN2Ly	$4.30\% \pm 2.64\%$
GF1Ly	$13.77\% \pm 6.87\%$
GF2Ly	$12.22\% \pm 7.89\%$

Table I: Classification error rates for model 'chair' in the test dataset. Average over 5 data realizations. The number of nodes is $N = 300$.

Stability experiment. We test the same trained Graph Neural Networks and Graph Filters with 2 layers on perturbed test point cloud models with different perturbation levels. We perturb the test point clouds by adding a Gaussian random variable with mean ϵ and variance 2ϵ to each coordinate of every sampling point, which can be seen as a deformation of the underlying manifold. We measure the stability by computing the difference between the error rates achieved based on the original test point cloud models and the perturbed ones. In Figure 5, we see that this difference increases when the perturbations become larger, but overall the differences are small. We also observe that Graph

Figure 3: Point cloud models with 300 sampling points in each model. Our goal is to identify chair models from other models such as toilet and table.

Neural Network is more stable compared with Graph Filters. Furthermore, the Graph Neural Networks and Graph Filters with Lipschitz continuous and integral Lipschitz continuous filters are more stable. Both of these observations validate our stability results.
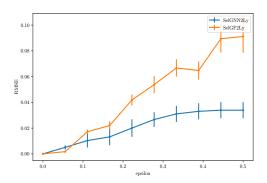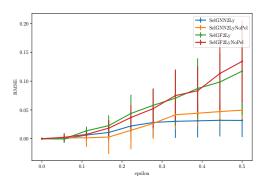

Figure 4: Difference between error rates on the original test dataset and the deformed one.

Figure 5: Difference between error rates on the original test dataset and the deformed one.

To further verify the discriminability under perturbations, we trained and tested the architectures with perturbed dataset. We can see from Table II that both GNN and GF can identify the chair model with small error rates while the error rates grow slightly with the increase of perturbation levels. GNNs still outperform GFs in discriminablity with the help of nonlinearity.

VIII. Conclusions

In this paper, we have defined manifold convolutions and manifold neural networks. We prove that the deformations on

Architecture	$\epsilon = 0.2$	0.4
GNN2Ly	$7.37\% \pm 1.43\%$	$7.71\% \pm 3.96\%$
GF2Ly	$13.76\% \pm 6.82\%$	$13.54\% \pm 7.16\%$
Architecture	$\epsilon = 0.6$	0.8
GNN2Ly	$8.04\% \pm 2.83\%$	$11.01\% \pm 6.33\%$
GF2Ly	$14.76\% \pm 5.67\%$	$16.04\% \pm 6.34\%$

Table II: Classification error rates for model 'chair' with perturbed training and test dataset. Average over 5 data realizations. The number of nodes is $N = 300$.

the embedded submanifolds can be represented as a form of perturbations to the Laplace-Beltrami operator. Considering the infinite dimensionality of LB operators, we import the definition of frequency difference threshold filters and frequency ratio threshold filters to help separate the spectrum. By assigning similar frequency responses to the eigenvalues that are close enough, these filters can be proved to be stable under absolute and relative perturbations to the LB operator respectively with Lipschitz continuous assumptions. While the manifold filters need to trade-off between the stability and discriminability. MNNs composed with layers of manifold filters and pointwise nonlinearities can be proved to be stable to absolute and relative perturbations to the LB operators. While the frequency mixing brought by pointwise nonlinearity can help with the discriminability. We conclude that the MNNs are thus both stable to deformations and discriminative. We also show the discretizations of MNNs in both spatial and time domains to make our proposed MNNs implementable. We finally verified our results numerically with a point cloud classification problem with ModelNet10 dataset.

References

[1] 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.

[2] ——, "Stability of neural networks on manifolds to relative perturbations," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5473–5477.

[3] 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.

[4] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 274–288, 2016.

[5] F. Gama, E. Isufi, G. Leus, and A. Ribeiro, "Graphs, convolutions, and neural networks: From graph filters to graph neural networks," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 128–138, 2020.

[6] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

[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] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[9] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *The World Wide Web Conference*, 2019, pp. 417–426.

[10] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 346–353.

[11] A. Chowdhury, G. Verma, C. Rao, A. Swami, and S. Segarra, "Unfolding wmmse using graph neural networks for efficient power allocation," *IEEE Transactions on Wireless Communications*, 2021.

[12] Z. Wang, M. Eisen, and A. Ribeiro, "Unsupervised learning for asynchronous resource allocation in ad-hoc wireless networks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8143–8147.

[13] 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.

[14] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," *arXiv preprint arXiv:2104.13478*, 2021.

[15] 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.

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

[17] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, G. M. Hernández *et al.*, *Signals & systems*. Pearson Educación, 1997.

[18] S. Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.

[19] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5680–5695, 2020.

[20] F. Gama, A. Ribeiro, and J. Bruna, "Stability of graph scattering transforms," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8038–8048, 2019.

[21] S. Verma and Z.-L. Zhang, "Stability and generalization of graph convolutional neural networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1539–1548.

[22] D. Zou and G. Lerman, "Graph convolutional neural networks via scattering," *Applied and Computational Harmonic Analysis*, vol. 49, no. 3, pp. 1046–1074, 2020.

[23] L. Ruiz, F. Gama, and A. Ribeiro, "Graph neural networks: Architectures, stability, and transferability," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 660–682, 2021.

[24] L. Ruiz, Z. Wang, and A. Ribeiro, "Graphon and graph neural network stability," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 5255–5259.

[25] S. Maskey, R. Levie, and G. Kutyniok, "Transferability of graph neural networks: an extended graphon approach," *arXiv preprint arXiv:2109.10096*, 2021.

[26] L. Ruiz, L. F. Chamon, and A. Ribeiro, "Transferability Properties of Graph Neural Networks," *arXiv preprint arXiv:2112.04629*, 2021.

[27] 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.

[28] 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.

[29] R. Levie, W. Huang, L. Bucci, M. Bronstein, and G. Kutyniok, "Transferability of spectral graph convolutional neural networks," *Journal of Machine Learning Research*, vol. 22, no. 272, pp. 1–59, 2021.

[30] J. Gallier and J. Quaintance, *Differential geometry and Lie groups: a computational perspective*. Springer Nature, 2020, vol. 12.

[31] Y. Canzani, *Analysis on manifolds via the Laplacian*. Lecture Notes available at: http://www. math. harvard. edu/canzani/docs/Laplacian. pdf, 2013, ch. 4, pp. 41–44.

[32] P. Moon and D. E. Spencer, *Field theory handbook: including coordinate systems, differential equations and their solutions*. Springer, 2012.

[33] 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.

[34] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[35] W. Arendt and W. P. Schleich, *Mathematical analysis of evolution, information, and complexity*. John Wiley & Sons, 2009.

[36] 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.

[37] 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.

[38] L. Lovász, *Large networks and graph limits*. American Mathematical Soc., 2012, vol. 60.

[39] D. B. Dunson, H.-T. Wu, and N. Wu, "Spectral convergence of graph laplacian and heat kernel reconstruction in l^∞ from random samples," *Applied and Computational Harmonic Analysis*, vol. 55, pp. 282–336, 2021.

[40] R. Merris, "A survey of graph laplacians," *Linear and Multilinear Algebra*, vol. 39, no. 1-2, pp. 19–31, 1995.

[41] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1912–1920.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[43] A. Seelmann, "Notes on the $sin2\theta$ theorem," *Integral Equations and Operator Theory*, vol. 79, no. 4, pp. 579–597, 2014.

[44] M. Belkin and P. Niyogi, "Convergence of laplacian eigenmaps," *Advances in neural information processing systems*, vol. 19, 2006.

[45] U. Von Luxburg, M. Belkin, and O. Bousquet, "Consistency of spectral clustering," *The Annals of Statistics*, pp. 555–586, 2008.

APPENDIX

A. Definition 1 and Convolutional Filters in Continuous Time

The manifold convolution in Definition 1 can also be motivated with a connection to linear time invariant filters. This requires that we consider the differential equation

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial}{\partial x} u(x,t). \tag{47}$$

This is a one-sided wave equation and it is therefore not an exact analogous of the diffusion equation in (6) – this would require that the second derivative be used in the right of (47). The important observation to make here is that the exponential of the derivative operator is a time shift so that we can write $u(x,t) = e^{t\partial/\partial x} f(x) = f(x-t)$. This is true because $e^{t\partial/\partial x} f(x)$ and $f(x-t)$ are both solutions of (47). It then follows that Definition 1 particularized to (47) yields the convolution definition

$$g(x) = \int_0^\infty \tilde{h}(t) e^{t\partial/\partial x} f(x)\, \mathrm{d}t. = \int_0^\infty \tilde{h}(t) f(x-t)\, \mathrm{d}t. \tag{48}$$

This is the standard definition of time convolutions.

The frequency representation result in Proposition 2 holds for (48) and it implies that standard convolutional filters in continuous time are completely characterized by the frequency response in Definition 2. The more standard definition of a filter's frequency response as the Fourier transform of the impulse response $\tilde{h}(t)$ – as opposed to the Laplace transform we use in Definition 2 – suffices because complex exponentials e^{jw} are an orthonormal basis of eigenfunctions of the derivative operator with associated eigenvalues $j\omega$.

B. Proof of Theorem 1

Based on equation (16) and the definition in (5), the operation carried out on the deformed manifold data f can be written as

$$-\mathcal{L}'f(x) = (\nabla \cdot \nabla)f(\tau(x)) \tag{49}$$
$$= (J(\tau_*)^T \nabla_\tau \cdot J(\tau_*)^T \nabla_\tau)f(\tau(x)). \tag{50}$$

The equality in (50) results from the chain rule of gradient operator where ∇_τ is denoted as the intrinsic gradient around $\tau(x)$ in the tangent space $T_{\tau(x)}\mathcal{M}$. By replacing $J(\tau_*) = I + \Delta$ the inner product term, (50) can be rewritten as

$$J(\tau_*)^T \nabla_\tau \cdot J(\tau_*)^T \nabla_\tau = \nabla_\tau \cdot \nabla_\tau + 2(\Delta^T \nabla_\tau \cdot \nabla_\tau) \\ + \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau. \tag{51}$$

With $\mathcal{L} = -\nabla_\tau \cdot \nabla_\tau$, the perturbed operator is

$$\mathcal{L} - \mathcal{L}' = 2(\Delta^T \nabla_\tau \cdot \nabla_\tau) + \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau \tag{52}$$
$$= 2\|\Delta\|_F(\nabla_\tau \cdot \nabla_\tau) + \|\Delta\|_F^2(\nabla_\tau \cdot \nabla_\tau) + \mathbf{A}. \tag{53}$$

From (52) to (53), we extract the relative term and use $\mathbf{A}$ to represent the compliment terms. This leads to $\mathcal{E} = \|\Delta\|_F^2 + 2\|\Delta\|_F$, as the relative perturbation term, the norm of which is bounded by the leading term as $O(\epsilon)$. The norm of the compliment term therefore can be written as

$$\|\mathbf{A}\| = \|\mathcal{E}(\nabla_\tau \cdot \nabla_\tau) - 2(\Delta^T \nabla_\tau \cdot \nabla_\tau) - \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau\| \tag{54}$$
$$\leq \|2\|\Delta\|_F(\nabla_\tau \cdot \nabla_\tau) - 2(\Delta^T \nabla_\tau \cdot \nabla_\tau)\| \\ + \|\|\Delta\|_F^2(\nabla_\tau \cdot \nabla_\tau) - \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau\|, \tag{55}$$

which can be also bounded by the leading terms as $O(\epsilon)$ combining with the boundedness of the gradient field.

C. Proof of Theorem 2

In the following, we denote $\langle \cdot, \cdot \rangle_{L^2(\mathcal{M})}$ as $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_{L^2(\mathcal{M})}$ as $\|\cdot\|$ for simplicity. We start by bounding the norm difference between two outputs of filter functions on operators $\mathcal{L}$ and $\mathcal{L}'$ defined in (15) as

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\| = \\ \left\|\sum_{i=1}^{\infty} \hat{h}(\lambda_i)\langle f, \phi_i\rangle\phi_i - \sum_{i=1}^{\infty} \hat{h}(\lambda_i')\langle f, \phi_i'\rangle\phi_i'\right\|. \tag{56}$$

We denote the index of partitions that contain a single eigenvalue as a set $\mathcal{K}_s$ and the rest as a set $\mathcal{K}_m$. We can decompose the filter function as $\hat{h}(\lambda) = h^{(0)}(\lambda) + \sum_{l \in \mathcal{K}_m} h^{(l)}(\lambda)$ with

$$h^{(0)}(\lambda) = \begin{cases} \hat{h}(\lambda) - \sum_{l \in \mathcal{K}_m} \hat{h}(C_l) & \lambda \in [\Lambda_k(\alpha)]_{k \in \mathcal{K}_s} \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

$$h^{(l)}(\lambda) = \begin{cases} \hat{h}(C_l) & \lambda \in [\Lambda_k(\alpha)]_{k \in \mathcal{K}_s} \\ \hat{h}(\lambda) & \lambda \in \Lambda_l(\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{58}$$

where C_l is some constant in $\Lambda_l(\alpha)$. We can start by analyzing the output difference of $h^{(0)}(\lambda)$. With the triangle inequality, the norm difference can then be written as

$$\left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i\rangle\phi_i - h^{(0)}(\lambda_i')\langle f, \phi_i'\rangle\phi_i'\right\|$$
$$= \left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i\rangle\phi_i - h^{(0)}(\lambda_i)\langle f, \phi_i'\rangle\phi_i' + \right. \\ \left. h^{(0)}(\lambda_i)\langle f, \phi_i'\rangle\phi_i' - h^{(0)}(\lambda_i')\langle f, \phi_i'\rangle\phi_i'\right\| \tag{59}$$
$$\leq \left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i\rangle\phi_i - h^{(0)}(\lambda_i)\langle f, \phi_i'\rangle\phi_i'\right\| + \\ \left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i'\rangle\phi_i' - h^{(0)}(\lambda_i')\langle f, \phi_i'\rangle\phi_i'\right\| \tag{60}$$
$$\leq \left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\left(\langle f, \phi_i\rangle\phi_i - \langle f, \phi_i\rangle\phi_i' + \langle f, \phi_i\rangle\phi_i' - \right. \right. \\ \left. \left. \langle f, \phi_i'\rangle\phi_i'\right)\right\| + \left\|\sum_{i=1}^{\infty}(h^{(0)}(\lambda_i) - h^{(0)}(\lambda_i'))\langle f, \phi_i'\rangle\phi_i'\right\| \tag{61}$$
$$\leq \left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i\rangle(\phi_i - \phi_i')\right\| \\ + \left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i - \phi_i'\rangle\phi_i'\right\| \\ + \left\|\sum_{i=1}^{\infty}(h^{(0)}(\lambda_i) - h^{(0)}(\lambda_i'))\langle f, \phi_i'\rangle\phi_i'\right\| \tag{62}$$

For the first term in (62), we employ Lemma 2 and therefore we have $\sigma = \lambda_i$ and $\omega = \lambda_i'$, for $\lambda_i \in [\Lambda_k(\alpha)]_{k \in \mathcal{K}_s}$ we can have

$$\|\phi_i - \phi_i'\| \leq \frac{\pi}{2}\frac{\|\mathbf{A}\|}{\alpha - \epsilon} = \frac{\pi}{2}\frac{\epsilon}{\alpha - \epsilon}. \tag{63}$$

Here d can be seen as $d = \min_{\lambda_i \in \Lambda_k(\alpha), \lambda_j \in \Lambda_l(\alpha), k \neq l}|\lambda_i - \lambda_j'|$. Combined with the fact that $|\lambda_i - \lambda_j| > \alpha$ and $|\lambda_i - \lambda_i'| \leq \epsilon$ for all $\lambda_i \in \Lambda_k(\alpha), \lambda_j \in \Lambda_l(\alpha), k \neq l$, we have $d \geq \alpha - \epsilon$. With Cauchy-Schwartz inequality, we have the first term in (62) bounded as

$$\left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i\rangle(\phi_i - \phi_i')\right\| \\ \leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i)||\langle f, \phi_i\rangle|\|\phi_i - \phi_i'\| \leq \frac{N_s \pi \epsilon}{2(\alpha - \epsilon)}\|f\|. \tag{64}$$

The second term in (62) is bounded as

$$\left\|\sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i - \phi_i'\rangle\phi_i'\right\| \\ \leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i)|\|\phi_i - \phi_i'\|\|f\| \leq \frac{N_s \pi \epsilon}{2(\alpha - \epsilon)}\|f\|. \tag{65}$$

These two bounds are obtained by noting that $|h^{(0)}(\lambda)| < 1$ and $h^{(0)}(\lambda) = 0$ for $\lambda \in [\Lambda_k(\alpha)]_{k \in \mathcal{K}_m}$. The number of eigenvalues within $[\Lambda_k(\alpha)]_{k \in \mathcal{K}_s}$ is denoted as N_s. The third term in (62)

can be bounded by the Lipschitz continuity of h combined with Lemma 1.

$$\left\| \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i) - h^{(0)}(\lambda_i'))\langle f, \phi_i'\rangle \phi_i' \right\|^2$$

$$\leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i) - h^{(0)}(\lambda_i')|^2 |\langle f, \phi_i'\rangle|^2$$

$$\leq \sum_{i=1}^{\infty} A_h^2 |\lambda_i - \lambda_i'|^2 |\langle f, \phi_i'\rangle|^2 \leq A_h^2 \epsilon^2 \|f\|^2. \tag{66}$$

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

$$\left\| \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathcal{L}')f \right\|$$

$$\leq \left\| (\hat{h}(C_l) + \delta)f - (\hat{h}(C_l) - \delta)f \right\| \leq 2\delta \|f\|, \tag{67}$$

where $\mathbf{h}^{(l)}(\mathcal{L})$ and $\mathbf{h}^{(l)}(\mathcal{L}')$ are manifold filters with filter function $h^{(l)}(\lambda)$ on the LB operators $\mathcal{L}$ and $\mathcal{L}'$ respectively. Combining the filter functions, we can write

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\| =$$

$$\left\| \mathbf{h}^{(0)}(\mathcal{L})f + \sum_{l \in \mathcal{K}_m} \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathcal{L}')f - \sum_{l \in \mathcal{K}_m} \mathbf{h}^{(l)}(\mathcal{L}')f \right\| \tag{68}$$

$$\leq \|\mathbf{h}^{(0)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathcal{L}')f\| + \sum_{l \in \mathcal{K}_m} \|\mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathcal{L}')f\| \tag{69}$$

$$\leq \frac{N_s \pi \epsilon}{\alpha - \epsilon} \|f\| + A_h \epsilon \|f\| + 2(N - N_s)\delta \|f\|, \tag{70}$$

which concludes the proof.

Supplemental Materials

A. Proof of Proposition 3

Weyl's law [35, Chapter 1] states that if $(\mathcal{M}, g)$ is a compact Riemannian manifold of dimension d, then

$$\lambda_k = \frac{C_1}{C_d^{2/d}} \left(\frac{k}{Vol(\mathcal{M})} \right)^{2/d}, \tag{71}$$

where C_d denotes the volume of the unit ball of $\mathbb{R}^d$ and C_1 is an arbitrary constant. Therefore, if $\lambda_{k+1} - \lambda_k \leq \alpha$, we have

$$(k+1)^{2/d} - k^{2/d} \leq \frac{\alpha}{C_1} (Vol(\mathcal{M})C_d)^{2/d} \tag{72}$$

while the left side can be scaled down to $(k+1)^{2/d} - k^{2/d} \geq \frac{2}{d} k^{2/d-1}$. This implies that

$$k^{\frac{2-d}{d}} \leq \left(\frac{\alpha d (Vol(\mathcal{M})C_d)^{2/d}}{2C_1} \right)^{\frac{d}{2-d}}, \tag{73}$$

with $d > 2$, we can claim that for

$$k > \lceil (\alpha d/C_1)^{d/(2-d)} (C_d Vol(\mathcal{M}))^{2/(2-d)} \rceil,$$

it holds that $\lambda_{k+1} - \lambda_k \leq \alpha$. Proof of Proposition 4 is similar and is also based on (71).

B. Proof of Theorem 3

The decomposition follows the same routine as (56) shows. By decomposing the filter function as (74) and (75), the norm difference can also be bounded separately.

$$h^{(0)}(\lambda) = \begin{cases} \hat{h}(\lambda) - \sum_{l \in \mathcal{K}_m} \hat{h}(C_l) & \lambda \in [\Lambda_k(\gamma)]_{k \in \mathcal{K}_s} \\ 0 & \text{otherwise} \end{cases} \tag{74}$$

$$h^{(l)}(\lambda) = \begin{cases} \hat{h}(C_l) & \lambda \in [\Lambda_k(\gamma)]_{k \in \mathcal{K}_s} \\ \hat{h}(\lambda) & \lambda \in \Lambda_l(\gamma) \\ 0 & \text{otherwise} \end{cases} \tag{75}$$

where now $\hat{h}(\lambda) = h^{(0)}(\lambda) + \sum_{l \in \mathcal{K}_m} h^{(l)}(\lambda)$ with $\mathcal{K}_s$ defined as the group index set of singletons and $\mathcal{K}_m$ the set of partitions that contain multiple eigenvalues. For manifold filter $\mathbf{h}^{(0)}(\mathcal{L})$ with filter function $h^{(0)}(\lambda)$, the norm difference can also be written as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i \rangle \phi_i - h^{(0)}(\lambda_i')\langle f, \phi_i' \rangle \phi_i' \right\|$$

$$\leq \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i \rangle (\phi_i - \phi_i') \right\|$$

$$+ \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i - \phi_i' \rangle \phi_i' \right\|$$

$$+ \left\| \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i) - h^{(0)}(\lambda_i'))\langle f, \phi_i' \rangle \phi_i' \right\|. \tag{76}$$

The difference of the eigenvalues due to relative perturbations can be similarly addressed by Lemma 3.

The first two terms of (76) rely on the differences of eigenfunctions, which can be derived with Davis-Kahan Theorem in Lemma 2, the difference of eigenfunctions can be written as

$$\|\mathcal{E}\mathcal{L}\phi_i\| = \|\mathcal{E}\lambda_i\phi_i\| = \lambda_i\|\mathcal{E}\phi_i\| \leq \lambda_i\|\mathcal{E}\|\|\phi_i\| \leq \lambda_i\epsilon. \tag{77}$$

The first term in (76) then can be bounded as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i \rangle (\phi_i - \phi_i') \right\|$$

$$\leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i)||\langle f, \phi_i \rangle| \|\phi_i - \phi_i'\| \leq \sum_{i \in \mathcal{K}_s} \frac{\pi \lambda_i \epsilon}{2d_i} \|f\|. \tag{78}$$

Because $d_i = \min\{|\lambda_i - \lambda_{i-1}'|, |\lambda_i' - \lambda_{i-1}|, |\lambda_{i+1}' - \lambda_i|, |\lambda_{i+1} - \lambda_i'|\}$, with Lemma 3 implied, we have

$$|\lambda_i - \lambda_{i-1}'| \geq |\lambda_i - (1+\epsilon)\lambda_{i-1}|, \tag{79}$$
$$|\lambda_i' - \lambda_{i-1}| \geq |(1-\epsilon)\lambda_i - \lambda_{i-1}|, \tag{80}$$
$$|\lambda_{i+1}' - \lambda_i| \geq |(1-\epsilon)\lambda_{i+1} - \lambda_i|, \tag{81}$$
$$|\lambda_{i+1} - \lambda_i'| \geq |\lambda_{i+1} - (1+\epsilon)\lambda_i|. \tag{82}$$

Combine with Lemma 3 and Definition 8, $d_i \geq \epsilon\gamma + \gamma - \epsilon$:

$$|(1-\epsilon)\lambda_{i+1} - \lambda_i| \geq |\gamma\lambda_i - \epsilon\lambda_{i+1}| \tag{83}$$

$$= \epsilon\lambda_i \left| 1 - \frac{\lambda_{i+1}}{\lambda_i} + \frac{\gamma}{\epsilon} - 1 \right| \tag{84}$$

$$\geq \lambda_i(\gamma - \epsilon + \gamma\epsilon) \tag{85}$$

This leads to the bound as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i \rangle (\phi_i - \phi_i') \right\| \leq \frac{M_s \pi \epsilon}{2(\gamma - \epsilon + \gamma\epsilon)} \|f\|. \tag{86}$$

The second term in (76) can also be bounded as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i - \phi_i' \rangle \phi_i' \right\|$$

$$\leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i)| \|\phi_i - \phi_i'\| \|f\| \leq \frac{M_s \pi \epsilon}{2(\gamma - \epsilon + \gamma\epsilon)} \|f\|, \tag{87}$$

which similarly results from the fact that $|h^{(0)}(\lambda)| < 1$ and $h^{(0)}(\lambda) = 0$ for $\lambda \in [\Lambda_k(\gamma)]_{k \in \mathcal{K}_m}$. The number of eigenvalues within $[\Lambda_k(\gamma)]_{k \in \mathcal{K}_s}$ is denoted as M_s.

The third term in (76) is:

$$\left\| \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i) - h^{(0)}(\lambda_i'))\langle f, \phi_i' \rangle \phi_i' \right\|^2$$

$$\leq \sum_{i=1}^{\infty} \left(\frac{B_h \epsilon |\lambda_i|}{(\lambda_i + \lambda_i')/2} \right)^2 \langle f, \phi_i' \rangle^2 \leq \left(\frac{2B_h\epsilon}{2-\epsilon} \right)^2 \|f\|^2, \tag{88}$$

with the use of Lemma 3 and Definition 10.

Then we need to analyze the output difference of $h^{(l)}(\lambda)$.

$$\left\| \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathcal{L}')f \right\|$$

$$\leq \left\| (\hat{h}(C_l) + \delta)f - (\hat{h}(C_l) - \delta)f \right\| \leq 2\delta\|f\|. \tag{89}$$

Combine the filter function, we could get

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\| =$$

$$\left\| \mathbf{h}^{(0)}(\mathcal{L})f + \sum_{l \in \mathcal{K}_m} \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathcal{L}')f - \sum_{l \in \mathcal{K}_m} \mathbf{h}^{(l)}(\mathcal{L}')f \right\|$$
(90)

$$\leq \|\mathbf{h}^{(0)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathcal{L}')f\| + \sum_{l \in \mathcal{K}_m} \|\mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathcal{L}')f\|$$
(91)

$$\leq \frac{M_s \pi \epsilon}{\gamma - \epsilon + \gamma \epsilon} \|f\| + \frac{2B_h \epsilon}{2 - \epsilon} \|f\| + 2(M - M_s)\delta \|f\|, \quad (92)$$

which concludes the proof.

B. Proof of Theorem 4

To bound the output difference of MNNs, we need to write in the form of features of the final layer

$$\|\Phi(\mathbf{H}, \mathcal{L}, f) - \Phi(\mathbf{H}, \mathcal{L}', f)\| = \left\| \sum_{q=1}^{F_L} f_L^q - \sum_{q=1}^{F_L} f_L'^q \right\|. \quad (93)$$

The output signal of layer l of MNN $\Phi(\mathbf{H}, \mathcal{L}, f)$ can be written as

$$f_l^p = \sigma \left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q \right). \quad (94)$$

Similarly, for the perturbed $\mathcal{L}'$ the corresponding MNN is $\Phi(\mathbf{H}, \mathcal{L}', f)$ the output signal can be written as

$$f_l'^p = \sigma \left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}'^q \right). \quad (95)$$

The difference therefore becomes

$$\|f_l^p - f_l'^p\|$$
$$= \left\| \sigma \left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q \right) - \sigma \left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}'^q \right) \right\|. \quad (96)$$

With the assumption that σ is normalized Lipschitz, we have

$$\|f_l^p - f_l'^p\| \leq \left\| \sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q - \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}'^q \right\| \quad (97)$$

$$\leq \sum_{q=1}^{F_{l-1}} \left\| \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q - \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}'^q \right\|. \quad (98)$$

By adding and subtracting $\mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}^q$ from each term, combined with the triangle inequality we can get

$$\left\| \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q - \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}'^q \right\|$$
$$\leq \left\| \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q - \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}^q \right\|$$
$$+ \left\| \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}^q - \mathbf{h}_l^{pq}(\mathcal{L}') f_{l-1}'^q \right\| \quad (99)$$

The first term can be bounded with (70) for absolute perturbations. The second term can be decomposed by Cauchy-Schwartz inequality and non-amplifying of the filter functions as

$$\left\| f_l^p - f_l'^p \right\| \leq \sum_{q=1}^{F_{l-1}} C_{per}\epsilon \|f_{l-1}^q\| + \sum_{q=1}^{F_{l-1}} \|f_{l-1}^q - f_{l-1}'^q\|, \quad (100)$$

where C_{per} representing the constant in the stability bound of manifold filters. To solve this recursion, we need to compute the bound for $\|f_l^p\|$. By normalized Lipschitz continuity of σ and the fact that $\sigma(0) = 0$, we can get

$$\|f_l^p\| \leq \left\| \sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q \right\| \leq \sum_{q=1}^{F_{l-1}} \|\mathbf{h}_l^{pq}(\mathcal{L})\| \|f_{l-1}^q\|$$

$$\leq \sum_{q=1}^{F_{l-1}} \|f_{l-1}^q\| \leq \prod_{l'=1}^{l-1} F_{l'} \sum_{q=1}^{F_0} \|f^q\|. \quad (101)$$

Insert this conclusion back to solve the recursion, we can get

$$\left\| f_l^p - f_l'^p \right\| \leq l C_{per}\epsilon \left(\prod_{l'=1}^{l-1} F_{l'} \right) \sum_{q=1}^{F_0} \|f^q\|. \quad (102)$$

Replace l with L we can obtain

$$\|\Phi(\mathbf{H}, \mathcal{L}, f) - \Phi(\mathbf{H}, \mathcal{L}', f)\|$$

$$\leq \sum_{q=1}^{F_L} \left(L C_{per}\epsilon \left(\prod_{l'=1}^{L-1} F_{l'} \right) \sum_{q=1}^{F_0} \|f^q\| \right). \quad (103)$$

With $F_0 = F_L = 1$ and $F_l = F$ for $1 \leq l \leq L - 1$, then we have

$$\|\Phi(\mathbf{H}, \mathcal{L}, f) - \Phi(\mathbf{H}, \mathcal{L}', f)\| \leq L F^{L-1} C_{per}\epsilon \|f\|, \quad (104)$$

which concludes the proof.

C. Lemmas and Propositions

Now we need to include two important lemmas to analyze the influence on eigenvalues and eigenfunctions caused by the perturbation.

Lemma 1 *[Weyl's Theorem] The eigenvalues of LB operators $\mathcal{L}$ and perturbed $\mathcal{L}' = \mathcal{L} + \mathbf{A}$ satisfy*

$$|\lambda_i - \lambda_i'| \leq \|\mathbf{A}\|, \text{ for all } i = 1, 2 \ldots \quad (105)$$

Proof of Lemma 1.

The minimax principle asserts that

$$\lambda_i(\mathcal{L}) = \max_{codim T = i-1} \lambda[\mathcal{L}, T] = \max_{codim T \leq i-1} \min_{u \in T, \|u\|=1} \langle \mathcal{L}u, u \rangle. \quad (106)$$

Then for any $1 \leq i$, we have

$$\lambda_i(\mathcal{L}') = \max_{codim T \leq i-1} \min_{u \in T, \|u\|=1} \langle (\mathcal{L} + \mathbf{A})u, u \rangle \quad (107)$$

$$= \max_{codim T \leq i-1} \min_{u \in T, \|u\|=1} (\langle \mathcal{L}u, u \rangle + \langle \mathbf{A}u, u \rangle) \quad (108)$$

$$\geq \max_{codim T \leq i-1} \min_{u \in T, \|u\|=1} \langle \mathcal{L}u, u \rangle + \lambda_1(\mathbf{A})) \quad (109)$$

$$= \lambda_1(\mathbf{A}) + \max_{codim T \leq i-1} \min_{u \in T, \|u\|=1} \langle \mathcal{L}u, u \rangle \quad (110)$$

$$= \lambda_i(\mathcal{L}) + \lambda_1(\mathbf{A}). \quad (111)$$

Similarly, we can have $\lambda_i(\mathcal{L}') \leq \lambda_i(\mathcal{L}) + \max_k \lambda_k(\mathbf{A})$. This leads to $\lambda_1(\mathbf{A}) \leq \lambda_i(\mathcal{L}') - \lambda_i(\mathcal{L}) \leq \max_k \lambda_k(\mathbf{A})$. This leads to the conclusion that:

$$|\lambda_i' - \lambda_i| \leq \|\mathbf{A}\|. \quad (112)$$

∎

To measure the difference of eigenfunctions, we introduce the Davis-Kahan $\sin\theta$ theorem as follows.

Lemma 2 (Davis-Kahan $\sin\theta$ Theorem) *Suppose the spectra of operators $\mathcal{L}$ and $\mathcal{L}'$ are partitioned as $\sigma\bigcup\Sigma$ and $\omega\bigcup\Omega$ respectively, with $\sigma\bigcap\Sigma = \emptyset$ and $\omega\bigcap\Omega = \emptyset$. Then we have*

$$\|E_{\mathcal{L}}(\sigma) - E_{\mathcal{L}'}(\omega)\| \leq \frac{\pi}{2}\frac{\|(\mathcal{L}' - \mathcal{L})E_{\mathcal{L}}(\sigma)\|}{d} \leq \frac{\pi}{2}\frac{\|\mathcal{L}' - \mathcal{L}\|}{d}, \tag{113}$$

where d satisfies $\min_{x\in\sigma, y\in\Omega}|x-y| \geq d$ and $\min_{x\in\Sigma, y\in\omega}|x-y| \geq d$.

Proof of Lemma 2. See [43]. ∎

Lemma 3 *The eigenvalues of LB operators $\mathcal{L}$ and perturbed $\mathcal{L}' = \mathcal{L} + \mathcal{EL}$ with $\|\mathcal{E}\| = \epsilon$ satisfy*

$$|\lambda_i - \lambda_i'| \leq \epsilon|\lambda_i|, \text{ for all } i = 1, 2\ldots \tag{114}$$

Proof of Lemma 3. With the assumption that $\mathcal{L}' = \mathcal{L} + \mathcal{EL}$, we have

$$\lambda_i(\mathcal{L} + \mathcal{EL}) = \max_{codim T \leq i-1} \min_{u\in T, \|u\|=1} \langle(\mathcal{L} + \mathcal{EL})u, u\rangle \tag{115}$$

$$= \max_{codim T \leq i-1} \min_{u\in T, \|u\|=1} (\langle\mathcal{L}u, u\rangle + \langle\mathcal{EL}u, u\rangle) \tag{116}$$

$$= \lambda_i(\mathcal{L}) + \max_{codim T \leq i-1} \min_{u\in T, \|u\|=1} \langle\mathcal{EL}u, u\rangle. \tag{117}$$

For the second term, we have

$$|\langle\mathcal{EL}u, u\rangle| \leq \langle|\mathcal{E}||\mathcal{L}|u, u\rangle \leq \epsilon\sum_i |\lambda_i(\mathcal{L})||\xi_i|^2 = \epsilon\langle|\mathcal{L}|u, u\rangle \tag{118}$$

Therefore, we have

$$\lambda_i(\mathcal{L} + \mathcal{EL}) \leq \lambda_i(\mathcal{L}) + \epsilon\max_{codim T \leq i-1} \min_{u\in T, \|u\|=1} \langle|\mathcal{L}|u, u\rangle$$
$$= \lambda_i(\mathcal{L}) + \epsilon|\lambda_i(\mathcal{L})|, \tag{119}$$
$$\lambda_i(\mathcal{L} + \mathcal{EL}) \geq \lambda_i(\mathcal{L}) - \epsilon|\lambda_i(\mathcal{L})|, \tag{120}$$
$$\lambda_i(\mathcal{L}) - \epsilon|\lambda_i(\mathcal{L})| \leq \lambda_i(\mathcal{L} + \mathcal{EL}) \leq \lambda_i(\mathcal{L}) + \epsilon|\lambda_i(\mathcal{L})|, \tag{121}$$

which concludes the proof. ∎

D. Proof of Proposition 6

Considering that the discrete points $\{x_1, x_2, \ldots, x_n\}$ are uniformly sampled from manifold $\mathcal{M}$ with measure μ, the empirical measure associated with $d\mu$ can be denoted as $p_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, where δ_{x_i} is the Dirac measure supported on x_i. Similar to the inner product defined in the $L^2(\mathcal{M})$ space (4), the inner product on $L^2(\mathbf{G}_n)$ is denoted as

$$\langle u, v\rangle_{L^2(\mathbf{G}_n)} = \int u(x)v(x)dp_n = \frac{1}{n}\sum_{i=1}^n u(x_i)v(x_i). \tag{122}$$

The norm in $L^2(\mathbf{G}_n)$ is therefore $\|u\|_{L^2(\mathbf{G}_n)}^2 = \langle u, u\rangle_{L^2(\mathbf{G}_n)}$, with $u, v \in L^2(\mathcal{M})$. For signals $\mathbf{u}, \mathbf{v} \in L^2(\mathbf{G}_n)$, the inner product is therefore $\langle\mathbf{u}, \mathbf{v}\rangle_{L^2(\mathbf{G}_n)} = \frac{1}{n}\sum_{i=1}^n [\mathbf{u}]_i[\mathbf{v}]_i$.

We first import the existing results from [44] which indicates the spectral convergence of the constructed Laplacian operator based on the graph $\mathbf{G}_n$ to the LB operator of the underlying manifold.

Theorem 6 (Theorem 2.1 [44]) *Let $X = \{x_1, x_2, \ldots x_n\}$ be a set of n points sampled i.i.d. from a d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$. Let $\mathbf{G}_n$ be a graph approximation of $\mathcal{M}$ constructed from X with weight values set as (37) with $t_n = n^{-1/(d+2+\alpha)}$ and $\alpha > 0$. Let $\mathbf{L}_n$ be the graph Laplacian of $\mathbf{G}_n$ and $\mathcal{L}$ be the Laplace-Beltrami operator of $\mathcal{M}$. Let λ_i^n be the i-th eigenvalue of $\mathbf{L}_n$ and ϕ_i^n be the corresponding normalized eigenfunction. Let λ_i and ϕ_i be the corresponding eigenvalue and eigenfunction of $\mathcal{L}$ respectively. Then, it holds that*

$$\lim_{n\to\infty} \lambda_i^n = \lambda_i, \quad \lim_{n\to\infty} |\phi_i^n(x_j) - \phi_i(x_j)| = 0, j = 1, 2\ldots, n \tag{123}$$

where the limits are taken in probability.

With the definitions of neural networks on graph $\mathbf{G}_n$ and manifold $\mathcal{M}$, the output difference can be written as

$$\|\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_n f) - \mathbf{P}_n\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f))\| = \left\|\sum_{q=1}^{F_L} \mathbf{x}_L^q - \sum_{q=1}^{F_L} \mathbf{P}_n f_L^q\right\|$$
$$\leq \sum_{q=1}^{F_L} \|\mathbf{x}_L^q - \mathbf{P}_n f_L^q\|. \tag{124}$$

By inserting the definitions, we have

$$\|\mathbf{x}_l^p - \mathbf{P}_n f_l^p\|$$
$$= \left\|\sigma\left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q\right) - \mathbf{P}_n\sigma\left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\right)\right\| \tag{125}$$

with $\mathbf{x}_0 = \mathbf{P}_n f$ as the input of the first layer. With a normalized Lipschitz nonlinearity, we have

$$\|\mathbf{x}_l^p - \mathbf{P}_n f_l^p\| \leq \left\|\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{P}_n\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\right\| \tag{126}$$

$$\leq \sum_{q=1}^{F_{l-1}} \left\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\right\| \tag{127}$$

The difference can be further decomposed as

$$\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\|$$
$$\leq \|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_n f_{l-1}^q$$
$$+ \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_n f_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\| \tag{128}$$
$$\leq \left\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_n f_{l-1}^q\right\|$$
$$+ \left\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_n f_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\right\| \tag{129}$$

The first term can be bounded as $\|\mathbf{x}_{l-1}^q - \mathbf{P}_n f_{l-1}^q\|$ with the initial condition $\|\mathbf{x}_0 - \mathbf{P}_n f_0\| = 0$. The second term can be denoted as D_{l-1}^n. With the iteration employed, we can have

$$\|\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_n f) - \mathbf{P}_n\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)\| \leq \sum_{l=0}^L \prod_{l'=l}^L F_{l'} D_l^n.$$

Therefore, we can focus on the difference term D_l^n, we omit the feature and layer index to work on a general form.

$$\|\mathbf{h}(\mathbf{L}_n)\mathbf{P}_n f - \mathbf{P}_n \mathbf{h}(\mathcal{L})f\|$$
$$\leq \left\| \sum_{i=1}^{\infty} \hat{h}(\lambda_i^n)\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n - \sum_{i=1}^{\infty} \hat{h}(\lambda_i)\langle f, \phi_i\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i \right\| \tag{130}$$

We decompose the α-FDT filter function as $\hat{h}(\lambda) = h^{(0)}(\lambda) + \sum_{l\in\mathcal{K}_m} h^{(l)}(\lambda)$ as equations (57) and (58) show. With the triangle inequality, we start by analyzing the output difference of $h^{(0)}(\lambda)$ as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i^n)\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n - \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)\langle f, \phi_i\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i \right\|$$
$$\leq \left\| \sum_{i=1}^{\infty} \left(h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i) \right) \langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n \right\|$$
$$+ \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) \left(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n - \langle f, \phi_i\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i \right) \right\|. \tag{131}$$

The first term in (131) can be bounded by leveraging the A_h-Lipschitz continuity of the frequency response. From the convergence in probability stated in (123), we can claim that for each eigenvalue $\lambda_i \leq \lambda_M$, for all $\epsilon_i > 0$ and all $\delta_i > 0$, there exists some N_i such that for all $n > N_i$, we have

$$\mathbb{P}(|\lambda_i^n - \lambda_i| \leq \epsilon_i) \geq 1 - \delta_i, \tag{132}$$

Letting $\epsilon_i < \epsilon$ with $\epsilon > 0$, with probability at least $\prod_{i=1}^{M}(1-\delta_i) := 1-\delta$, the first term is bounded as

$$\left\| \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i))\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n \right\|$$
$$\leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i)| |\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}| \|\phi_i^n\| \tag{133}$$
$$\leq \sum_{i=1}^{N_s} A_h|\lambda_i^n - \lambda_i| \|\mathbf{P}_n f\| \|\phi_i^n\|^2 \leq N_s A_h \epsilon, \tag{134}$$

for all $n > \max_i N_i := N$.

The second term in (131) can be bounded combined with the convergence of eigenfunctions in (136) as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) \left(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n - \langle f, \phi_i\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i \right) \right\|$$
$$\leq \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) \left(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n - \langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\mathbf{P}_n\phi_i \right) \right\|$$
$$+ \left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) \left(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\mathbf{P}_n\phi_i - \langle f, \phi_i\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i \right) \right\| \tag{135}$$

From the convergence stated in (123), we can claim that for some fixed eigenfunction ϕ_i, for all $\epsilon_i > 0$ and all $\delta_i > 0$, there exists some N_i such that for all $n > N_i$, we have

$$\mathbb{P}(|\phi_i^n(x_j) - \phi_i(x_j)| \leq \epsilon_i) \geq 1 - \delta_i, \quad \text{for all } x_j \in X. \tag{136}$$

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 (135) can be bounded as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i) \left(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\phi_i^n - \langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i \right) \right\|$$
$$\leq \sum_{i=1}^{N_s} \|\mathbf{P}_n f\| \|\phi_i^n - \mathbf{P}_n\phi_i\| \leq N_s\epsilon, \tag{137}$$

because the frequency response is non-amplifying as stated in Assumption 1. The last equation comes from the definition of norm in $L^2(\mathbf{G}_n)$. The second term in (135) can be written as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i^n)(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n}\mathbf{P}_n\phi_i - \langle f, \phi_i\rangle_{\mathcal{M}}\mathbf{P}_n\phi_i) \right\|$$
$$\leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i^n)| |\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n} - \langle f, \phi_i\rangle_{\mathcal{M}}| \|\mathbf{P}_n\phi_i\|. \tag{138}$$

Because $\{x_1, x_2, \cdots, x_n\}$ is a set of uniform sampled points from $\mathcal{M}$, based on Theorem 19 in [45] we can claim that there exists some N such that for all $n > N$

$$\mathbb{P}(|\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n} - \langle f, \phi_i\rangle_{\mathcal{M}}| \leq \epsilon) \geq 1-\delta, \tag{139}$$

for all $\epsilon > 0$ and $\delta > 0$. 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 for all $\epsilon > 0$ and $\delta > 0$,

$$\mathbb{P}\left(\left\| \sum_{i=1}^{M} \hat{h}(\lambda_i^n) \left(\langle \mathbf{P}_n f, \phi_i^n\rangle_{\mathbf{G}_n} - \langle f, \phi_i\rangle_{\mathcal{M}} \right) \mathbf{P}_n\phi_i \right\| \leq M\epsilon \right)$$
$$\geq 1-\delta, \tag{140}$$

for all $n > N$.

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

$$\left\| \mathbf{P}_n \mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathbf{L}_n)\mathbf{P}_n f \right\|$$
$$\leq \left\| (\hat{h}(C_l) + \delta)\mathbf{P}_n f - (\hat{h}(C_l) - \delta)\mathbf{P}_n f \right\| \leq 2\delta\|\mathbf{P}_n f\|, \tag{141}$$

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

$$\|\mathbf{P}_n\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathbf{L}_n)\mathbf{P}_n f\|$$
$$= \left\| \mathbf{P}_n\mathbf{h}^{(0)}(\mathcal{L})f + \mathbf{P}_n \sum_{l\in\mathcal{K}_m} \mathbf{h}^{(l)}(\mathcal{L})f - \right.$$
$$\left. \mathbf{h}^{(0)}(\mathbf{L}_n)\mathbf{P}_n f - \sum_{l\in\mathcal{K}_m} \mathbf{h}^{(l)}(\mathbf{L}_n)\mathbf{P}f \right\| \tag{142}$$
$$\leq \|\mathbf{P}_n\mathbf{h}^{(0)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathbf{L}_n)\mathbf{P}_n f\| +$$
$$\sum_{l\in\mathcal{K}_m} \|\mathbf{P}_n\mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathbf{L}_n)\mathbf{P}_n f\|. \tag{143}$$

Above all, we can claim that there exists some N, such that for all $n > N$, for all $\epsilon' > 0$ and $\delta > 0$, we have

$$\mathbb{P}(\|\mathbf{h}(\mathbf{L}_n)\mathbf{P}_n f - \mathbf{P}_n\mathbf{h}(\mathcal{L})f\| \leq \epsilon') \geq 1-\delta. \tag{144}$$

With $\lim_{n\to\infty} D_l^n = 0$ in high probability, this concludes the proof.