Stability to Deformations in Manifold Neural Networks

Zhiyang Wang  Luana Ruiz  Alejandro Ribeiro

Abstract—Stability is an important property of graph neural networks (GNNs) which explains their success in many problems of practical interest. Existing GNN stability results depend on the size of the graph, restricting applicability to graphs of moderate size. To understand the stability properties of GNNs on large graphs, we define manifold convolutions and consider neural networks supported on manifolds. These are defined in terms of manifold diffusions mediated by the Laplace-Beltrami (LB) operator and are interpreted as limits of GNNs running on graphs of growing size. We define manifold deformations and show that they lead to perturbations of the manifold’s LB operator that consist of an absolute and a relative perturbation term. We then define two frequency dependent manifold filters that split the infinite dimensional spectrum of the LB operator in finite partitions, and prove that these filters are stable to absolute and relative perturbations of the LB operator respectively. We also observe a trade-off between the stability and the discriminability from the stability bounds. Moreover, manifold neural networks (MNNs) composed of these filters inherit the stability properties while the nonlinear activation function helps to improve the discriminability. Therefore, the MNNs can be both stable and discriminative. We verify our results numerically in shape classification with point cloud datasets.

Index Terms—Manifold convolution, manifold neural network, stability analysis, manifold deformation

I. INTRODUCTION

Inspired by convolutional neural networks (CNNs), graph neural networks (GNNs) are made up of layers consisting of banks of graph convolutional filters and pointwise nonlinearities [3]–[5]. A growing body of work has demonstrated the success of GNNs in tasks as diverse as learning ratings in recommendation systems [6], [7] and learning resource allocations in wireless communication networks [8], [9]. Recent theoretical work has demonstrated that the good performance of GNNs is explained by equivariance and stability properties they inherit from convolutions [10]. This is akin to seminal work proving that the same is true of CNNs [11]. Where CNNs can be shown to be equivariant to translation and stable to deformations of Euclidean space [11], GNNs can be shown to be equivariant to permutations and stable to deformations of the graph [12], [13].

Existing stability analyses of GNNs yield bounds that grow with the number of nodes in the graph [12], [13]. This is not only an important limitation but one that runs counter to empirical and theoretical evidence: GNNs are known to work well for large graphs and we can think of Euclidean space as a dense limit of a discrete graph. Our goal in this paper is to fill this gap by studying the stability properties of manifold neural networks (MNNs). A manifold is a geometric object with possibly irregular structure. Thus, discretizations of a manifold yield models for possibly irregular graphs of arbitrarily large size with well defined limits [14]–[16]. In studying stability of MNNs, we derive results that hold for such limit objects and therefore illustrate the stability of GNNs for graphs representing discretizations of the manifold with arbitrarily many nodes.

The importance of stability to deformations in understanding the merits of generic CNNs relative to the respective convolutional filters traces back to the seminal work on Euclidean scattering transforms [11]. This work inspired extensions of the analysis to graph scattering transforms [13], [17] and GNNs [12], [13], [18]. Common to all of these works is the introduction of domain (graph, or Euclidean) deformations and the proof of Lipschitz stability with respect to the norm of the deformation for some filter categories. The similar stability properties of CNNs and GNNs have been shown to be a consequence of their shared algebraic structure [19].

All of the GNN stability results described above have bounds that grow with the number of nodes in the graph. Overcoming this limitation has motivated the study of neural networks on graphons [10], [20], [21]. Results in this setting are independent of graph size—same as the results presented here, they hold for a limit object—but it must be pointed out that graphons are limits of dense graphs in the sense that they have growing degree. This is different from graphs sampled from manifolds that have constant degree as the discretization becomes finer [16]. Of particular relevance to our paper is work on GNN transferability for graphs that are sampled from a manifold [16]. This work shows that GNNs on a graph sampled from a manifold are close to the respective MNN. They can be combined with our stability results to prove stability of GNNs on graphs sampled from perturbed manifolds.

We start by proposing the definition of the manifold convolution and constructing MNNs with this operation. To study the stability properties of MNNs, we first show that manifold deformations lead to absolute and relative perturbations of the manifold’s Laplace-Beltrami (LB) operator (Theorem 2). Since perturbations of the LB operator perturb its spectrum, we introduce two types of manifold convolutional filters, the frequency difference threshold (FDT) filter (Definition 5) and the frequency ratio threshold (FRT) filter (Definition 9). These filters mitigate the spectral effects of absolute and relative eigenvalue perturbations respectively, grouping eigenvalues that are less than a certain threshold apart to make sure that
they have similar frequency responses. When satisfying certain Lipschitz continuity assumptions (Definitions 6 and 10), it is possible to show that these manifold filters are stable to absolute and relative perturbations of the LB operator respectively, because they mitigate the difference of frequency responses (Theorems 3 and 4). Using FDT and FRT filters along with Lipschitz and integral Lipschitz, we assume that

\[ M \ni x, x' \rightarrow \text{dist}(x, x') \leq \mu \]

between these points. This deformation satisfies\( \text{dist}(x, x') = \epsilon + 2 \Delta \) with \( \| \Delta \|_F = \epsilon \). Then, under mild assumptions,

\[ \| \Phi(f \circ \tau) - \Phi(f) \|_{L^2(M)} = O(\epsilon) \| f \|_{L^2(M)}. \]

This is the main result of our paper. It is worth noting that the Lipschitz continuity assumptions on the FDT and FRT filters inevitably result in a trade-off between their stability and discriminability. This means that manifold filters cannot be both stable and discriminative. Since MNNs inherit their stability properties from manifold filters, we could expect them to also inherit this incompatibility. Interestingly, they do not, because pointwise nonlinearities improve discriminability by scattering frequency components all over the eigenvalue spectrum. MNNs are hence as stable as but more discriminative than manifold filters or, alternatively, as discriminative as but more stable than manifold filters. These considerations are discussed in detail in Section III-C. Using GNNs—as we discuss in Section IV—MNNs are implemented as GNNs in practice by discretization of the manifold and time domains—we further illustrate this result numerically in Section V.

The rest of this paper is organized as follows. Section II-F defines manifold convolution and proves the stability of the proposed manifold filters. The stability analysis of MNNs is presented in Section III. Section IV illustrates the discretization of the MNN in the spatial and time domains. Section V describes the numerical experiments. Concluding remarks are presented in Section VI and proofs are deferred to the supplementary material.

II. STABILITY OF MANIFOLD FILTERS

Consider a compact, smooth and differentiable \( d \)-dimensional submanifold \( M \subset \mathbb{R}^N \). For simplicity, in this paper we use the words submanifold and manifold interchangeably, assuming the manifold \( M \) to always be embedded in \( \mathbb{R}^N \). This embedding induces a Riemannian structure \[ [22]. \] In turn, the Riemannian structure allows defining a measure \( \mu \) over the manifold as well as a notion of length for smooth curves on \( M \). Given two points \( x, y \in M \), the length of the shortest curve between \( x \) and \( y \) is denoted \( \text{dist}(x,y) \) and called the geodesic distance between these points.

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

\[ \langle f, g \rangle_{L^2(M)} = \int_M f(x)g(x)d\mu(x) \]

where \( d\mu(x) \) is the \( d \)-dimensional volume element corresponding to measure \( \mu \). Thus, the energy of the signal \( f \) is given by

\[ \| f \|_{L^2(M)}^2 = \langle f, f \rangle_{L^2(M)}. \]

A. Laplace-Beltrami Operator

On manifolds, differentiation is implemented by the intrinsic gradient, which is a local operator acting on a neighborhood of each point \( x \in M \) that is homeomorphic to a \( d \)-dimensional Euclidean space. This neighborhood, called the tangent space of \( x \in M \), contains all of the vectors which are tangent to \( M \) at \( x \) and is denoted as \( T_x M \). The disjoint union of all tangent spaces on \( M \) is the tangent bundle \( TM \). Formally, the intrinsic gradient is the operator \( \nabla : L^2(M) \rightarrow L^2(TM) \) mapping scalar functions \( f \in L^2(M) \) to tangent vector functions \( \nabla f(x) \in L^2(T_x M) \), which indicate the direction of the fastest change of the signal \( f \) at point \( x \). The adjoint of the intrinsic gradient is the intrinsic divergence, defined as \( \text{div} : L^2(TM) \rightarrow L^2(M) \). 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 \[ [14]. \]

The Laplace-Beltrami (LB) operator of a manifold \( M \) is defined as the operator \( \mathcal{L} : L^2(M) \rightarrow L^2(M) \) given by the function composition of the intrinsic divergence and the intrinsic gradient \[ [23]. \]

\[ \mathcal{L}f = -\text{div} \circ \nabla f = -\nabla \cdot \nabla f. \]

Much like the Laplace operator in Euclidean domains (or the Laplacian matrix, in the case of graphs \[ [24] \]), the LB operator measures the total variation of the 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 \[ [14]. \] Since \( \mathcal{L} \), like \( \nabla \), is a local operator depending on the tangent space \( T_x M \) of each point \( x \in M \), 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 the study of partial differential equations (PDEs), as it governs the dynamics of the diffusion of heat over manifolds as given by the heat equation

\[ \frac{\partial u(x,t)}{\partial t} + \mathcal{L} u(x,t) = 0. \]

If \( u(x,t) \in L^2(M) \) measures the temperature of point \( x \in M \) at time \( t \in \mathbb{R}^+ \), \[ (3) \] 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

\[ u(x,t) = e^{-t\mathcal{L}} f(x), \]

\[ (4) \]
we need to substitute the solution of the heat equation (4) for \( \phi \) where

\[
\phi = \int_0^\infty \phi_i e^{-\lambda_i t} dt,
\]

which is the key support to implement the LB operator in the following.

Because \( \mathcal{M} \) is compact, \( \mathcal{L} \) is a self-adjoint and positive-semidefinite operator. As such, the LB operator has a real positive eigenvalue spectrum \( \{ \lambda_i \}_{i=1}^\infty \) which satisfies

\[
\mathcal{L} \phi_i = \lambda_i \phi_i
\]

where \( \phi_i \) is the eigenfunction associated with eigenvalue \( \lambda_i \).

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 \( \lambda_i \propto i^{2/d} \) as a consequence of Weyl’s law [25]. The eigenfunctions \( \phi_i \) 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 \( \lambda_i \) can be interpreted as the canonical frequencies and the eigenfunctions \( \phi_i \) as the canonical oscillation modes of \( \mathcal{M} \).

\section{Manifold Filters}

Time signals are processed by filters which compute the continuous time convolution of the input signal and the filter impulse response [26]. Images and high-dimensional Euclidean signals are processed by filters implementing multidimensional convolutions [27], and graph signals are filtered by computing graph convolutions [28]. In this paper, we define a manifold filter as the convolution of the filter impulse response \( \tilde{h} \) and the manifold signal \( f \). Note that the definition of the convolution operation, denoted \( \ast_{\mathcal{M}} \), leverages the heat equation dynamics described in [4].

\begin{definition}[Manifold filter] Let \( \tilde{h} : \mathbb{R}^+ \to \mathbb{R} \) and let \( f \in L^2(\mathcal{M}) \) be a manifold signal. The manifold filter with impulse response \( \tilde{h} \), denoted \( h \), is given by

\[
g(x) = (h f)(x) := (\tilde{h} \ast_{\mathcal{M}} f)(x) := \int_0^\infty \tilde{h}(t) u(x, t) dt
\]

where \( \tilde{h} \ast_{\mathcal{M}} f \) is the manifold convolution of \( \tilde{h} \) and \( f \), and \( u(x, t) \) is the solution of the heat equation [3] with initial condition \( u(x, 0) = f(x) \).
\end{definition}

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 [3] starting at \( f(x) \) by \( h(t) \) 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 [4] for \( u(x, t) \) in (6). This leads to a closed-form expression of \( h \) that is parametric on the LB operator as shown in Proposition [1].

\begin{proposition}
The manifold filter \( h \) is a parametric map \( h(\mathcal{L}) \) of the LB operator \( \mathcal{L} \),

\[
g(x) = (h f)(x) = \int_0^\infty \tilde{h}(t) e^{-\lambda \cdot t^L} f(x) dt = h(\mathcal{L}) f(x)
\]

Proposition [1] tells us that manifold filters are spatial operators, because 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 \( e^{-\lambda \cdot t^L} \) can be seen as a diffusion or shift akin to a time delay in a linear time-invariant (LTI) filter [20].

Or, as the graph shift operator in a linear shift-invariant (LSI) graph filter [28]. Indeed, if we consider the manifold \( \mathcal{M} \) to be the real line, \( \mathcal{L} \) recovers a LTI filter. If we consider it to be a set of points connected by a geometric graph, \( \mathcal{L} \) recovers a LSI graph filter. We discuss these special cases in further detail in Section IV and Appendix A.

\section{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 \( \{ \phi_i \} \) as

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

where we say that \( \hat{f} \) is the frequency representation of the signal.

Frequency representations are useful because they help understand the frequency behavior of the manifold filter \( h(\mathcal{L}) \). To see this, consider the frequency representation of the manifold filter output \( g \) in (7).

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

Rearranging the integrals and substituting \( e^{-\lambda \cdot t^L} \phi_i = e^{-\lambda_\ell} \phi_i \), we get

\[
[\hat{g}]_i = \int_0^\infty \tilde{h}(t) e^{-\lambda_\ell} dt \hat{f}_i.
\]

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

\begin{definition}[Frequency response]
The frequency response of the filter \( h(\mathcal{L}) \) is given by

\[
\hat{h}(\lambda) = \int_0^\infty \tilde{h}(t) e^{-\lambda t} dt.
\]

An important consequence of Definition 2 is that, since \( \hat{h}(\lambda) \) is parametric on \( \lambda \), the manifold filter is pointwise in the frequency domain. This can be seen by plugging (11) into (10), and is stated explicitly in Proposition 2.

\begin{proposition}
The manifold filter \( h(\mathcal{L}) \) is pointwise in the frequency domain,

\[
[\hat{g}]_i = \hat{h}(\lambda_i)[\hat{f}]_i
\]

Proposition 2 also emphasizes that the frequency response of a manifold filter is independent of the underlying manifold. Note that, in (11), \( \hat{h}(\lambda) \) is a function of an arbitrary scalar variable \( \lambda \). 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 \( \lambda_i \) [cf. 5]. 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 \( \lambda_i' \), i.e., at the eigenvalues of the new LB operator \( \mathcal{L}' \).
D. Manifold Deformations

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

Let \( f: \mathbb{R} \rightarrow \mathbb{R} \) be a manifold signal. Because \( M \) is the codomain of \( \tau \), \( g = f \circ \tau \) maps points \( \tau(x) \in M \) 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 \( p = Lg \). Since \( p \) is also a signal on \( M \), we may define an operator \( L' \) mapping \( f \) directly into \( p \),

\[
p(x) = L'f(x) = Lg(x) = Lf(\tau(x)).
\]

The operator \( L' \) is the perturbed LB operator, which is effectively the new LB operator resulting from the deformation \( \tau \). Assuming that the gradient field is smooth, the difference between \( L' \) and \( L \) is given by the following theorem. The proof is deferred to Appendix C.

**Theorem 2** Let \( L \) be the LB operator of the manifold \( M \). Let \( \tau(x): M \rightarrow M \) be a manifold perturbation such that \( \text{dist}(x, \tau(x)) = \epsilon \) and \( J(\tau_x) = I + \Delta \) with \( \|\Delta\|_F = \epsilon \). If the gradient field is smooth, it holds that

\[
L - L' = EL + A,
\]

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

Therefore, the perturbation of the LB operator incurred by a manifold deformation \( \tau \) is a combination of an absolute perturbation \( A \) [cf. Definition 3] and a relative perturbation \( EL \) [cf. Definition 7]. This largely simplifies our analysis of stability. Since manifold filters are parametric on \( L \) [cf. Proposition 3], to characterize their stability to deformations of the manifold it suffices to analyze their behavior in the presence of absolute and relative LB perturbations. This is what we do in Sections II-E and II-F.

E. 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 \( L \) be the LB operator of manifold \( M \). An absolute perturbation of \( L \) is defined as

\[
L' = L + A,
\]

where the absolute perturbation operator \( A \) is symmetric.

Like \( L \), the operator \( L' \) resulting from the absolute perturbation of \( L \) is self-adjoint due to the symmetry of \( A \). Hence, it admits an eigendecomposition similar to (2). When the filter coefficients are fixed, the frequency response of the manifold filter [12] can be obtained by evaluating \( h(\lambda) \) at each \( \lambda_i \). 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 \( \lambda_i \). The challenge in this case is that the spectrum of \( L \) 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 manifold \( M \) and let \( L \) be its LB operator with eigenvalues \( \{\lambda_i\}_{i=1}^{\infty} \). Let \( C_d \) be an arbitrary constant and \( C_d \) the volume of the \( d \)-dimensional unit ball. For any \( \alpha > 0 \) and \( d > 2 \), there exists \( N_1 \),

\[
N_1 = \left[ (\alpha d/C_d)^{d/(2-d)}(C_d \text{Vol}(M))^{2/(2-d)} \right]
\]

such that, for all \( k > N_1 \),

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

**Proof.** This is a direct consequence of Weyl’s law [25].

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 filter called Frequency Difference Threshold (FDT) filter as introduced in Definition 5.

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

\[
|\lambda_i - \lambda_j| > \alpha.
\]

**Definition 5 (\( \alpha \)-FDT filter)** The \( \alpha \)-frequency difference threshold (\( \alpha \)-FDT) filter is defined as a filter \( h(\lambda) \) whose frequency response satisfies

\[
|h(\lambda_i) - h(\lambda_j)| \leq \delta_k \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\alpha)
\]

with \( \delta_k \leq \delta \) for \( k = 1, \ldots, N \).
In the \(\alpha\)-separated spectrum, eigenvalues \(\lambda_i \in \Lambda_k(\alpha)\) and \(\lambda_j \in \Lambda_1(\alpha)\) in different sets \((k \neq l)\) are at least \(\alpha\) away from each other. Conversely, eigenvalues \(\lambda_i, \lambda_j \in \Lambda_k(\alpha)\) are no more than \(\alpha\) apart. This partitioning creates several eigenvalue groups spaced by at least \(\alpha\). 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 \(\alpha\)-FDT filter. This filter separates the spectrum of the manifold by assigning similar frequency responses—that deviate no more than \(\delta_k\) from each other—to eigenvalues \(\lambda_i \in \Lambda_k(\alpha), 1 \leq k \leq N\). In other words, the \(\alpha\)-FDT filter does not discriminate between eigenvalues \(\lambda_i, \lambda_j \in \Lambda_k(\alpha)\). Importantly, the \(\delta_k\) in Definition 4 are finite, so that they can be bounded by some \(\delta\).

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

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

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

(19)

Between the eigenvalue groups, the filters that we consider are thus Lipschitz continuous with Lipschitz constant \(A_h\). 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 (11) that are both Lipschitz and \(\alpha\)-FDT.

Under mild assumptions on the amplitude of the filters \(h\) (Assumption 1), it can be shown that Lipschitz continuous \(\alpha\)-FDT filters are stable to absolute perturbations of the LB operator. This result is stated in Theorem 3.

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

Note that this assumption is rather reasonable, because the filter function \(h(\lambda)\) can always be normalized.

**Theorem 3 (Manifold filter stability to absolute perturbations)** Consider a manifold \(\mathcal{M}\) with LB operator \(\mathcal{L}\). Let \(h(\mathcal{L})\) be a \(\alpha\)-FDT manifold filter [cf. Definition 3] and \(A_h\)-Lipschitz [cf. Definition 6]. Consider an absolute perturbation \(\mathcal{L}' = \mathcal{L} + A\) of the LB operator \(\mathcal{L}\) [cf. Definition 3] where \(\|A\| = \epsilon < \alpha\). Then, under Assumption 1 it holds that

\[
\|h(\mathcal{L})f - h(\mathcal{L}')f\|_{L^2(\mathcal{M})} \leq \left(\frac{\pi N \alpha \epsilon}{\alpha - \epsilon} + A_h \epsilon + 2(N - N_\alpha)\delta\right) \|f\|_{L^2(\mathcal{M})},
\]

(20)

where \(N\) is the size of the \(\alpha\)-separated spectrum partition [cf. Definition 4] and \(N_\alpha\) is the number of singletons.

**Proof.** See appendix C. \(\blacksquare\)

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 \(\alpha\), 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 fluctuation of the filter frequency response within the same eigenvalue group.

The bound in Theorem 3 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 3 it holds that

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

(21)

A particular case of Theorem 3 is 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 \(\alpha\) is large. On the other hand, small \(A_h\) and large \(\alpha\) 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. Importantly, this trade-off is not related to the magnitude of the frequencies that the filters amplify (as is the case in, e.g., [12], [20]). 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 III-C.

**F. Stability of Manifold Filters to Relative Perturbations**

Relative perturbations of the LB operator are defined 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

\[
\mathcal{L}' - \mathcal{L} = E \mathcal{L},
\]

(22)

where the relative perturbation operator \(E\) is symmetric.

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 \(\alpha\)-close to each other to different groups. 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 M be a d-dimensional 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
\[
N_2 = \lceil (C_1 (\gamma + 1)^{d/2} - 1)^{-1} \rceil
\]
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 \[25\].

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

Definition 8 (\( \gamma \)-separated spectrum.) The \( \gamma \)-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
\[
\left| \frac{\lambda_i}{\lambda_j} - 1 \right| > \gamma.
\]

Definition 9 (\( \gamma \)-FRT filter.) The \( \gamma \)-frequency ratio threshold \( \gamma \)-FRT filter is defined as a filter \( \hat{h}(\mathcal{L}) \) whose frequency response satisfies
\[
|\hat{h}(\lambda_i) - \hat{h}(\lambda_j)| \leq \delta_k, \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\gamma)
\]
with \( \delta_k \leq \delta \) for \( k = 1, 2, \ldots, M \).

In the \( \gamma \)-separated spectrum, the sets \( \Lambda_k(\gamma) \) are built based on eigenvalue distances relative to the eigenvalues’ magnitudes and weighted by the parameter \( \gamma \). 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 \( \gamma \)-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.

For a manifold filter to be stable to relative perturbations of the LB operator, we need a further restriction on their variability. 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. Instead, we need our filters to be integral Lipschitz as described in Definition 10.

Definition 10 (Integral Lipschitz filter) A filter is integral Lipschitz with constant \( B_h \) if its frequency response is given by
\[
|\hat{h}(a) - \hat{h}(b)| \leq \frac{B_h |a - b|}{(a + b)/2} \text{ for all } a, b \in (0, \infty).
\]

Figure 2: Illustration of a \( \gamma \)-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.

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

Under Assumption 1 integral Lipschitz \( \gamma \)-FRT filters are stable to relative perturbations as stated in Theorem 5.

Theorem 4 (Manifold filter stability to relative perturbations) Consider a manifold \( M \) with LB operator \( \mathcal{L} \). Let \( \hat{h}(\mathcal{L}) \) be a \( \gamma \)-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 \( \mathcal{L}' = \mathcal{L} + EC \) of the LB operator \( \mathcal{L} \) [cf. Definition 7] where \( ||E|| = \epsilon < \gamma \). Then, under Assumption 1 it holds that
\[
||\hat{h}(\mathcal{L}) \mathcal{L} - \hat{h}(\mathcal{L}') \mathcal{L}||_{L^2(M)} \leq \left( \frac{\pi M \epsilon}{\gamma - \epsilon + \gamma \epsilon} + \frac{2 B_h \epsilon}{2 - \epsilon} \right) ||f||_{L^2(M)}
\]
where \( M \) is the size of the \( \gamma \)-separated spectrum partition [cf. Definition 8].

Proof. See Appendix D.

When \( \epsilon \) is sufficiently small (\( \epsilon \ll \min(\gamma, 2) \)), which is typically the case with deformations such as the one in Theorem 2, the denominators on the right hand side of \( (27) \) are approximately equal to \( \gamma \) and 2 respectively. Hence, \( \gamma \)-FRT integral Lipschitz filters are stable to relative perturbations of the LB operator. Besides appearing in the bound in Theorem 4, the frequency ratio threshold \( \gamma \) also affects stability indirectly through the partition size \( M \). With a larger \( \gamma \), 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 \( \gamma \) make for smoother filters which in turn lead to a less discriminative manifold filter. Therefore, integral Lipschitz \( \gamma \)-FRT filters also exhibit a trade-off between discriminability and stability.

Remark 1 By comparing the illustrations of \( \alpha \)-FDT filter (Definition 5) and \( \gamma \)-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 effect on the low-frequency components. In the low frequencies, the eigengaps $\lambda_{i+1} - \lambda_i$ may be smaller than the difference threshold $\alpha$, but larger than the relative ratio threshold $\gamma \lambda_i$ due to $\lambda_i$ being small. However, for appropriate values of $\gamma$ a filter may be both FDT and FRT. This will be shown in Section III.

III. 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

$$g_p^l(x) = \sum_{q=1}^{F_{l-1}} h_{pq}^l(L) f_{l-1}^q(x), \quad (28)$$

where $h_{pq}^l(L)$ is the filter mapping the $q$-th feature from layer $l-1$ to the $p$-th feature of layer $l$ as in (12), for $1 \leq q \leq F_{l-1}$ and $1 \leq p \leq F_l$. The intermediate features are then processed by a nonlinearity $\sigma: \mathbb{R} \rightarrow \mathbb{R}$ as

$$f_l^p(x) = \sigma(g_p^l(x)). \quad (29)$$

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

Assumption 2 (Normalized Lipschitz activation functions)
The activation function $\sigma$ is normalized Lipschitz continuous, i.e., $|\sigma(a) - \sigma(b)| \leq |a - b|$, 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 $f_0^q$ for $1 \leq q \leq F_0$. At the output of the MNN, the output features are given by the outputs of the $L$-th layer, i.e., $f_L^q$ for $1 \leq q \leq F_L$. To represent the MNN more succinctly, we may gather the impulse responses of the manifold convolutional filters $h_{pq}^l(L)$ across all layers in a function set $H$, and define the MNN map $\Phi(H, L, f)$. This map emphasizes that the MNN is parameterized by both the filter functions and the LB operator $L$. We next will analyze the stability of $\Phi(H, L, f)$ 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 5 (MNN stability) Consider a manifold $M$ with LB operator $L$. Let $\Phi(H, L, f)$ be an $L$-layer MNN on $M$ 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 $h(L)$ and nonlinearity functions satisfy Assumptions 7 and 2 respectively. Let $L'$ 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 $\|h(L) f - h(L') f\|_{L^2(M)} \leq C_{per} \epsilon \|f\|_{L^2(M)}$, it holds that $\|\Phi(H, L, f) - \Phi(H, L', f)\|_{L^2(M)} \leq L F^{L-1} C_{per} \epsilon \|f\|_{L^2(M)}$.

Proof. The conclusions follow directly from Theorem 5 combined with Theorem 3 or Theorem 7 under the corresponding assumptions.

Combining Theorem 5 with Proposition 5, we observe that $\alpha$-FDT manifold filters with Lipschitz continuity can be composed to construct MNNs which are stable to absolute perturbations; while $\gamma$-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 (29), we see that other than the perturbation size $\epsilon$, there are three terms that determine the stability of MNNs. The first term is $L F^{L-1}$, 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, \( \alpha_0 \) 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 \( C_{pert} \) brings along the trade-off between stability and discriminability. However, unlike manifold filters, MNNs can be both stable and discriminative. This is because of the nonlinear activation function, as we discuss in further detail in Section III-C.

B. Stability of MNNs to Manifold Deformations

In Theorem 5 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 \( \tau(x) : M \to M \), 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 \( \alpha \)-FDT and \( \gamma \)-FRT, and both Lipschitz continuous and integral Lipschitz continuous. The spectrum can be made to be both \( \alpha \)-separated and \( \gamma \)-separated by making sure the eigenvalues in different partitions satisfy both (17) and (24). 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 6** Let \( M \) be a manifold with LB operator \( \mathcal{L} \) and \( f \) be manifold data. We construct \( \Phi(H, L, f) \) as a MNN on \( M \) where the filters \( h(L) \) are \( \alpha \)-FDT [cf. Definition 3], \( \alpha / \lambda_1 \)-FRT [cf. Definition 9], \( \alpha_0 \)-Lipschitz [cf. Definition 7] and \( B_h \)-integral Lipschitz [cf. Definition 10]. Consider a deformation on \( M \) as \( \tau(x) : M \to M \) 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 2 and 3 it holds that

\[
\|\Phi(H, L, f) - \Phi(H, L', f)\|_{L^2(M)} = O(\epsilon)\|f\|_{L^2(M)}.
\] (32)

Together, Theorems 2-5 imply that MNNs are stable to the manifold deformations \( \nu \) 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 \( \alpha \)-FDT [cf. Definition 3], \( \gamma \)-FRT [cf. Definition 9], Lipschitz [cf. Definition 7] and integral Lipschitz [cf. Definition 10]. We can propose an easier special case to relate \( \alpha \) and \( \gamma \) by utilizing the spectrum property of LB operator. By setting the \( \alpha \)-FDT filter with \( \alpha = \gamma \lambda_1 \), eigenvalues \( \lambda_i, \lambda_{i+1} \in \Lambda_k(\alpha) \) would lead to \( \lambda_i, \lambda_{i+1} \in \Lambda_k(\gamma) \) due to the fact that

\[
\lambda_{i+1} - \lambda_i \leq \alpha = \gamma \lambda_1 \leq \gamma \lambda_i,
\] (33)

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

C. Discussion

**Stability vs. discriminability tradeoff.** In both stability theorems for manifold filters (Theorems 3 and 4) and in the stability theorem for MNNs (Theorem 5), the stability bounds depend on the frequency partition threshold (\( \alpha \) or \( \gamma \)), the number of total partitions (\( N \) or \( M \)) and the Lipschitz continuity constant (\( \alpha_0 \) or \( B_h \)). The frequency partition threshold and the number of partitions have a combined effect on stability. As indicated by Definitions 3 and 8, a larger frequency threshold leads to a smaller number of partitions (\( \Lambda_k(\alpha) \) or \( \Lambda_k(\gamma) \)). This decreases the total number of partitions (\( N \) or \( M \)), as eigenvalues that would otherwise be separated for small thresholds end up being grouped when the threshold is large.

While a large frequency threshold makes for a larger number of singletons, the total number of partitions either stays the same or decreases because the number of eigenvalues does not exceed the number of partitions [cf. Proposition 3]. Thus, a larger frequency threshold and a smaller number of partitions both lead to a smaller stability bound. Simultaneously, a large frequency threshold makes it so that the spectrum is 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 (\( \alpha_0 \) 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 tends 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 [12], which can be see as instantiations of MNNs on discrete samples of the manifold as further discussed in Section IV.

**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 [10], [20], [29], [30]. Manifolds are however more powerful, because they can represent the limit of graphs with both bounded and unbounded degrees [31]—the graphon is only the limit of sequences of dense graphs [32]. 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 [20], [33] 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, on manifolds 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.

IV. PRACTICAL IMPLMENTATION OF MANIFOLD 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 the architecture described by [25] in applications. In this section, we discuss how MNNs are implemented in practice over a discrete sample of 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 sample points. From these points’ coordinates, the structure of the manifold is approximated by a geometric or a nearest neighbor graph [31], [34], [35]. The LB operator is then approximated by the graph Laplacian, which can be shown to converge to the LB operator as the number of sampled points grows [34] [35].

Explicitly, suppose that \( X = \{x_1, x_2, \ldots, x_n\} \) is a set of \( n \) points sampled i.i.d. from measure \( \mu \) of manifold \( \mathcal{M} \subset \mathbb{R}^N \), which is embedded in \( \mathbb{R}^N \). We can construct a complete weighted symmetric graph \( G_n \) 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

\[
w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{4t_n}\right),
\]

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 [31]. The adjacency matrix \( A_n \in \mathbb{R}^{n \times n} \) is thus defined as \( A_{n_{ij}} = w_{ij} \) for \( 1 \leq i, j \leq n \) and the corresponding graph Laplacian matrix \( L_n \) [36] is given by

\[
L_n = \text{diag}(A_n)1 - A_n.
\]

We interpret \( L_n \) the Laplacian operator of the discretized manifold \( G_n \). Similarly, to sample manifold signals we construct a uniform sampling operator \( P_n : L^2(M) \rightarrow L^2(X) \). Given a manifold signal \( f \), we can use \( P_n \) to sample graph signals \( x_n \in \mathbb{R}^n \) as

\[
x_n = P_nf \text{ with } |x_n|_1 = f(x_i), \quad x_i \in X
\]

where the \( i \)-th entry of the graph signal \( x_n \) is the manifold signal \( f \) evaluated at the sample point \( x_i \).

In Proposition [4] we have shown that the manifold filter \( h \) is parametric on the LB operator. Therefore, we can also parameterize \( h \) by the discrete Laplacian operator \( L_n \), i.e.,

\[
z_n = \int_0^\infty \hat{h}(t)e^{-tL_n}dtx_n = h(L_n)x_n, \quad x_n, z_n \in \mathbb{R}^n,
\]

where \( z_n \), the output of the filter, is now a discrete graph signal. By cascading these discrete manifold filters and pointwise nonlinearities layer after layer, we can then approximate the MNN on this discretized manifold as

\[
x_n^l = \sigma \left( \sum_{q=1}^{F_l} h^{pq}_l(L_n)x_{n-1}^q \right),
\]

where \( h^{pq}_l(L) \) 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 \( x_n^l \) and \( x_{n-1}^l \) for simplicity). Further gathering the filter functions in the set \( H \), this neural network can be represented more succinctly as \( \Phi(H, L_n, x) \).

Equation (38) is a consistent approximation of the MNN because, as \( n \) goes to infinity, the discrete graph Laplacian operator \( L_n \) of the discretized manifold \( G_n \) converges to the LB operator \( L \) of the manifold \( M \), and the sampled graph signal \( x_n \) converges to the manifold signal \( f \) [31]. Together, these facts imply that the output of the neural network on the discretized manifold converges to the output of the neural network on the continuous manifold as stated in the following proposition.

Proposition 6 Let \( X = \{x_1, x_2, \ldots, x_n\} \) be \( n \) points sampled i.i.d. from measure \( \mu \) of manifold \( \mathcal{M} \subset \mathbb{R}^N \), with corresponding sampling operator \( P_n \) [35]. Let \( G_n \) be a discrete approximation of \( M \) constructed from \( X \) as in [34] with \( t_n = n^{-1/(d+2+\alpha)} \) and \( \alpha > 0 \). Let \( \Phi(H, L_n, \cdot) \) be a neural network parametrized either by the LB operator \( L \) of the manifold \( M \) [29] or the discrete Laplacian operator \( L_n \) of the discretized manifold \( G_n \). With probability one, it holds that

\[
\lim_{n \rightarrow \infty} \|\Phi(H, L_n, P_n f) - P_nf, \Phi(H, L, f)\| = 0.
\]

Proof. See Appendix F. \( \blacksquare \)

This proposition provides theoretical support to claim that neural networks constructed from the discrete Laplacian \( L \) inherit the stability properties of the MNN.

B. Discretization in the Time Domain

In order to learn an MNN [29], we need to learn the manifold convolutional filters \( h^{pq}_l \). This means that we need to learn
Neural network architectures. We build dense graphs to approximate the point cloud models. By connecting a point with all the other points in the point cloud, the edge weight is defined based on 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 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.

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

| Architecture | error rates |
|--------------|-------------|
| GNN1Ly       | 8.04\% ± 0.88\% |
| GNN2Ly       | 4.39\% ± 2.04\% |
| GF1Ly        | 13.77\% ± 6.87\% |
| GF2Ly        | 12.22\% ± 7.89\% |

Dataset. We evaluate our MNN stability results on the ModelNet10 [27] 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.

V. NUMERICAL EXPERIMENTS

Combining (37) and (41), 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,

$$
\hat{h}(L) x = \sum_{k=0}^{K-1} h_k e^{-k L} x.
$$

Equation (43) recovers the definition of the graph convolution [28] with graph shift operator $e^{-L}$. 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 V.

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 $\epsilon$ and variance $\nu$ 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 [3] we see that this difference increases when the perturbations become larger, but overall the differences are small. We also observe that Graph 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.

To further verify the discriminability under perturbations, we trained and tested the architectures with perturbed dataset. We can see from Table I that both GNN and GF can identify the chair model with small error rates while the error rates grow.
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.

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.

slightly with the increase of perturbation levels. GNNs still outperform GFs in discriminability with the help of nonlinearity.

| Architecture | $\epsilon = 0.2$ | 0.4 |
|--------------|-----------------|-----|
| GNN2Ly       | 7.37% ± 1.43%   | 7.71% ± 3.96% |
| GF2Ly        | 13.76% ± 6.82%  | 13.54% ± 7.16% |

| Architecture | $\epsilon = 0.6$ | 0.8 |
|--------------|-----------------|-----|
| GNN2Ly       | 8.04% ± 2.83%   | 11.01% ± 6.33% |
| GF2Ly        | 14.76% ± 5.97%  | 16.04% ± 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 $N = 300$.

VI. CONCLUSIONS

In this paper, we have defined manifold convolutions and manifold neural networks. We prove that the deformations on 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,” arXiv preprint arXiv:2110.04702, 2021.
[3] 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.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
The motivation of formalizing the manifold convolutions in Definition [1] can be seen more directly by connecting LTI filter with wave diffusions in 1-dimensional space. With initial condition set as $u(x, 0) = f(x)$ and $\frac{\partial u(x, t)}{\partial t} = 0$ for $x \in \mathbb{R}$, the diffusion of wave is given by

$$
\frac{\partial^2 u(x, t)}{\partial t^2} = \mathcal{L} u(x, t).
$$

The solution to this equation is given by $u(x, t) = e^{t\sqrt{\mathcal{L}}} f(x) + e^{-t\sqrt{\mathcal{L}}} f(x)$. With Taylor expansion employed to the exponential terms and $\sqrt{\mathcal{L}} = \frac{\partial}{\partial x}$, we can get the conclusion that $u(x, t) = f(x-t) + f(x + t)$ [39]. This indicates that the wave condition at point $x \in \mathbb{R}$ at time $t$ is actually a translation of the initial wave condition with a time period $t$. We can define the convolution over this 1-d space with $f \in L^2(\mathbb{R})$ passing through a filter with impulse response $\bar{h} : \mathbb{R}^+ \to \mathbb{R}$ as

$$
g(x) = \int_0^\infty \bar{h}(t) f(x-t) dt + \int_{-\infty}^0 \bar{h}(-t) f(x-t) dt,
$$

which can be further written as

$$
g(x) = \int_0^\infty \bar{h}(t) (f(x-t) + f(x + t)) dt = \int_0^\infty \bar{h}(t) u(x, t) dt.
$$

The final expression shows exactly the same form as Definition [1] which supports our underlying meaning of the manifold convolution definition. We can also understand the definition of manifold convolution in higher dimensional space from the view of graph convolutions. In Section [IV] we also show how graph neural networks can be recovered by sampling graphs from the manifold $\mathcal{M}$.

**B. Proof of Theorem [2]**

Based on equation [15] and the definition of Laplace operator in [2], the operation carried out on the deformed manifold data $f$ can be written as

$$
-L^2 f(x) = (\nabla \cdot \nabla) f(\tau(x))
$$

$$
= (J(\tau_x) \nabla \cdot J(\tau_x)^T \nabla) f(\tau(x)).
$$

The equality in (47) results from the chain rule of gradient operator where $\nabla_\tau$ is denoted as the intrinsic gradient around $\tau(x)$ in the tangent space $T_{\tau(x)}\mathcal{M}$. By replacing $J(\tau_x) = I + \Delta$ the inner product term in (47) can be rewritten as

$$
J(\tau_x) \nabla_\tau \cdot J(\tau_x)^T \nabla_\tau = \nabla_\tau \cdot \nabla_\tau + 2(\Delta^T \nabla_\tau \cdot \nabla_\tau + \Delta^T \nabla_\tau \cdot \nabla_\tau).
$$

APPENDIX

**A. Remark of Definition [7]**
With $\mathcal{L} = -\nabla_\tau \cdot \nabla_\tau$, the perturbed operator can be further written as

$$\mathcal{L} - \mathcal{L}' = 2(\Delta^T \nabla_\tau \cdot \nabla_\tau) + \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau$$

(50)

$$= 2\|\Delta\|_F(\nabla_\tau \cdot \nabla_\tau) + \|\Delta\|_F^2(\nabla_\tau \cdot \nabla_\tau) + A.$$  

(51)

From (50) to (51), we extract the relative term and use $A_1$ to represent the complement terms. This leads to

$$E = \|\Delta\|^2_2 + 2\|\Delta\|_F,$$  

(52)

as the relative perturbation term, the norm of which is bounded by the leading term as $O(\epsilon)$. The norm of the complement term therefore can be written as

$$\|A\| = \|E(\nabla_\tau \cdot \nabla_\tau) - 2(\Delta^T \nabla_\tau \cdot \nabla_\tau) - \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau\|_F$$

(53)

$$\leq \|2\Delta\|_F(\nabla_\tau \cdot \nabla_\tau) - 2(\Delta^T \nabla_\tau \cdot \nabla_\tau)\|_F$$

$$+ \|\Delta\|_F^2(\nabla_\tau \cdot \nabla_\tau) - \Delta^T \nabla_\tau \cdot \Delta^T \nabla_\tau\|_F,$$

(54)

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

C. Proof of Theorem 3

In the following, we denote $\langle \cdot, \cdot \rangle_{L^2(M)}$ as $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_{L^2(M)}$ 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 (12) as

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

(55)

We denote the index of partitions that contain a single eigenvalue as a set $K_s$ and the rest as a set $K_m$. We can decompose the filter function as $h(\lambda) = h^{(0)}(\lambda) + \sum_{\lambda \in K_m} h^{(1)}(\lambda)$ with

$$h^{(0)}(\lambda) = \left\{ \begin{array}{ll} h(\lambda) - \sum_{\lambda \in K_s} h(C_i) & \lambda \in [\Lambda_s(\alpha)]_{k \in K_s} \\ 0 & \text{otherwise} \end{array} \right.$$  

(56)

$$h^{(1)}(\lambda) = \left\{ \begin{array}{ll} h(C_i) & \lambda \in [\Lambda_s(\alpha)]_{k \in K_s} \\ h(\lambda) & \lambda \in [\Lambda_s(\alpha)]_{k \in K_m} \\ 0 & \text{otherwise} \end{array} \right.$$  

(57)

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

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

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

(58)

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

(59)

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

(60)

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

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

(61)

For the first term in (61), we employ Lemma 2 and therefore we have $\sigma = \lambda_i$ and $\omega = \lambda_i'$ for $\lambda_i \in [\Lambda_s(\alpha)]_{k \in K_s}$, we can have

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

(62)

Here $d$ can be seen as $d = \min_{\lambda_i \in [\Lambda_s(\alpha)], \lambda_j \in [\Lambda_s(\alpha), k \neq l]} |\lambda_i - \lambda_j|$. Combined with the fact that $|\lambda_i - \lambda_j| > \alpha$ and $|\lambda_i - \lambda_j| \leq \epsilon$ for all $\lambda_i \in K_s(\alpha), \lambda_j \in \Lambda_s(\alpha), k \neq l$, we have $d \geq \alpha - \epsilon$. With Cauchy-Schwartz inequality, we have the first term in (61) bounded as

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

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

(63)

The second term in (61) is bounded as

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

(64)

These two bounds are obtained by noting that $|h^{(0)}(\lambda)| < 1$ and $h^{(0)}(\lambda) = 0$ for $\lambda \in [\Lambda_s(\alpha)]_{k \in K_m}$. The number of eigenvalues within $[\Lambda_s(\alpha)]_{k \in K_m}$ is denoted as $N_s$. The third term in (61)
can be bounded by the Lipschitz continuity of $h$ combined with Lemma [1]

$$\left\| \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i) - h^{(0)}(\lambda'_i)) (f, \phi'_i) \phi'_i \right\|^{2}$$

$$\leq \sum_{i=1}^{\infty} |h^{(0)}(\lambda_i) - h^{(0)}(\lambda'_i)|^{2} |(f, \phi'_i)|^{2}$$

$$\leq \sum_{i=1}^{\infty} A_{h}^{2} |\lambda_i - \lambda'_i|^{2} |(f, \phi'_i)|^{2} \leq A_{h}^{2} \epsilon^{2} \|f\|^{2}. \quad (65)$$

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

$$\left\| h^{(i)}(L)f - h^{(i)}(L')f \right\|$$

$$\leq \|(h(C_l) + \delta)f - (h(C_l) - \delta)f\| \leq 2\delta\|f\|, \quad (66)$$

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

$$\left\| h(L)f - h(L')f \right\|$$

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

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

$$\leq N_{s}\pi \epsilon \|f\| + A_{h}\epsilon \|f\| + 2(N - N_{s})\delta \|f\|, \quad (67)$$

$$\leq N_{s}\pi \epsilon \|f\| + A_{h}\epsilon \|f\| + 2(N - N_{s})\delta \|f\|, \quad (68)$$

$$\leq \frac{N_{s}\pi \epsilon}{\alpha - \epsilon} \|f\| + A_{h}\epsilon \|f\| + 2(N - N_{s})\delta \|f\|, \quad (69)$$

which concludes the proof.

**D. Proof of Theorem 4**

The decomposition follows the same routine as [55] shows. By decomposing the filter function as (70) and (71), the norm difference can also be bounded separately.

$$h^{(0)}(\lambda) = \begin{cases} h(\lambda) - \sum_{i \in \mathcal{K}_m} h(C_l), & \lambda \in [\Lambda_{k}(\gamma)]_{k \in \mathcal{K}_s} \\ 0, & \text{otherwise} \end{cases} \quad (70)$$

$$h^{(i)}(\lambda) = \begin{cases} h(C_l), & \lambda \in [\Lambda_{k}(\gamma)]_{k \in \mathcal{K}_s} \\ h(\lambda), & \lambda \in [\Lambda_{i}(\gamma)]_{k \in \mathcal{K}_s} \\ 0, & \text{otherwise} \end{cases} \quad (71)$$

where now $h(\lambda) = h^{(0)}(\lambda) + \sum_{i \in \mathcal{K}_m} h^{(i)}(\lambda)$ 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 $h^{(0)}(L)$ with filter function $h^{(0)}(\lambda)$, the norm difference can also be written as

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

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

$$+ \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)(f, \phi_i - \phi'_i)\phi'_i$$

$$+ \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i) - h^{(0)}(\lambda'_i))(f, \phi'_i)\phi'_i. \quad (72)$$

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

The first two terms of (72) 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

$$\| h^{(i)}(\lambda) \| = \| \lambda \| \| \phi_i \| \leq \lambda_i \| \epsilon \| \| \phi_i \| \leq \lambda_i \epsilon. \quad (73)$$

The first term in (72) then can be bounded as

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

$$\leq \sum_{i=1}^{\infty} \left\| h^{(0)}(\lambda_i)(f, \phi_i) \right\| \| \phi_i - \phi'_i \| \leq \sum_{i \in \mathcal{K}_s} \frac{\pi \lambda_i \epsilon}{2d_i} \|f\|. \quad (74)$$

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

$$|\lambda_i - \lambda'_{i-1}| \geq |\lambda_i - (1 + \epsilon)\lambda_{i-1}|, \quad (75)$$

$$|\lambda'_i - \lambda_{i-1}| \geq |(1 - \epsilon)\lambda_{i-1} - \lambda'_i|, \quad (76)$$

$$|\lambda'_{i+1} - \lambda_i| \geq |(1 - \epsilon)\lambda_{i+1} - \lambda'_i|, \quad (77)$$

$$|\lambda_{i+1} - \lambda'_i| \geq |\lambda_{i+1} - (1 + \epsilon)\lambda_i|. \quad (78)$$

Combine with Lemma [3] and Definition [8] $d_i \geq \epsilon \gamma + \gamma - \epsilon: \quad (79)$

$$\left| (1 - \epsilon)\lambda_{i+1} - \lambda'_i \right| \geq |\gamma\lambda_i - \epsilon\lambda_{i+1}|$$

$$= \epsilon\lambda_i \left| 1 - \frac{\lambda_{i+1}}{\lambda_i} + \frac{\gamma}{\epsilon} - 1 \right| \geq \lambda_i (\gamma - \epsilon + \gamma) \quad (80)$$

This leads to the bound as

$$\left\| \sum_{i=1}^{\infty} h^{(0)}(\lambda_i)(f, \phi_i)(\phi_i - \phi'_i) \right\| \leq \frac{M_{s}\pi \epsilon}{2(\gamma - \epsilon + \gamma)} \|f\|. \quad (82)$$

The second term in (72) also can be bounded as

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

$$\leq \sum_{i=1}^{\infty} \left\| h^{(0)}(\lambda_i)\right\| \| \phi_i - \phi'_i \| \|f\| \leq \frac{M_{s}\pi \epsilon}{2(\gamma - \epsilon + \gamma)} \|f\|. \quad (83)$$

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 (72) is:
\[
\left\| \sum_{i=1}^{\infty} (h^{(0)}(\lambda_i) - h^{(0)}(\lambda'_i)) \langle f, \Phi_i \rangle \Phi_i \right\|^2 \\
\leq \sum_{i=1}^{\infty} \left( \frac{B_h \epsilon(\lambda_i)}{(\lambda_i + \lambda'_i)/2} \right)^2 \langle f, \Phi_i \rangle^2 \leq \left( \frac{2B_h \epsilon}{2 - \epsilon} \right)^2 \|f\|^2,
\]
with the use of Lemma 3 and Definition 10.

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

\[
\left\| h^{(l)}(\mathcal{L}) f - h^{(l)}(\mathcal{L}') f \right\| \\
\leq \| (h(C_l + \delta) f - (h(C_l) - \delta) f \| \leq 2\delta \| f \|,
\]

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

\[
\left\| h^{(l)}(\mathcal{L}) f - h^{(l)}(\mathcal{L}') f \right\| = \\
\left\| h^{(l)}(\mathcal{L}) f + \sum_{l \in \mathcal{K}_m} h^{(l)}(\mathcal{L}) f - h^{(0)}(\mathcal{L}') f - \sum_{l \in \mathcal{K}_m} h^{(l)}(\mathcal{L}') f \right\|
\leq \| h^{(0)}(\mathcal{L}) f - h^{(0)}(\mathcal{L}') f \| + \sum_{l \in \mathcal{K}_m} \| h^{(l)}(\mathcal{L}) f - h^{(l)}(\mathcal{L}') f \|
\leq \frac{M_\lambda \pi \epsilon}{\gamma - \epsilon + \gamma \epsilon} \| f \| + \frac{2B_h \epsilon}{2 - \epsilon} \| f \| + 2(M - M_\lambda) \delta \| f \|,
\]

which concludes the proof.

\[E. \ Proof \ of \ Theorem \ 5\]

To bound the output difference of MNNs, we need to write in the form of features of the final layer
\[
\left\| \Phi(H, \mathcal{L}, f) - \Phi(H, \mathcal{L}', f) \right\| = \left\| \sum_{q=1}^{F_l} f^q_l - \sum_{q=1}^{F_l} f^q_l \right\|.
\]

The output signal of layer \( l \) of MNN \( \Phi(H, \mathcal{L}, f) \) can be written as
\[
f^p_l = \sigma \left( \sum_{q=1}^{F_l-1} h^{pq}_l(\mathcal{L}) f^q_{l-1} \right).
\]

Similarly, for the perturbed \( \mathcal{L}' \) the corresponding MNN is \( \Phi(H, \mathcal{L}', f) \) the output signal can be written as
\[
f'^p_l = \sigma \left( \sum_{q=1}^{F_l-1} h^{pq}_l(\mathcal{L}') f'^q_{l-1} \right).
\]

The difference therefore becomes
\[
\left\| f^p_l - f'^p_l \right\| = \left\| \sigma \left( \sum_{q=1}^{F_l-1} h^{pq}_l(\mathcal{L}) f^q_{l-1} \right) - \sigma \left( \sum_{q=1}^{F_l-1} h^{pq}_l(\mathcal{L}') f'^q_{l-1} \right) \right\|.
\]

With the assumption that \( \sigma \) is normalized Lipschitz, we have
\[
\left\| f^p_l - f'^p_l \right\| \leq \left\| \sum_{q=1}^{F_l-1} h^{pq}_l(\mathcal{L}) f^q_{l-1} - h^{pq}_l(\mathcal{L}') f'^q_{l-1} \right\|
\leq \sum_{q=1}^{F_l-1} \left\| h^{pq}_l(\mathcal{L}) f^q_{l-1} - h^{pq}_l(\mathcal{L}') f'^q_{l-1} \right\|.
\]

By adding and subtracting \( h^{pq}_l(\mathcal{L}') f^q_{l-1} \) from each term, combined with the triangle inequality we can get
\[
\left\| h^{pq}_l(\mathcal{L}) f^q_{l-1} - h^{pq}_l(\mathcal{L}') f^q_{l-1} \right\|
\leq \left\| h^{pq}_l(\mathcal{L}) f^q_{l-1} - h^{pq}_l(\mathcal{L}') f^q_{l-1} \right\|
+ \left\| h^{pq}_l(\mathcal{L}') f^q_{l-1} - h^{pq}_l(\mathcal{L}') f'^q_{l-1} \right\|.
\]

The first term can be bounded with (99) for absolute perturbations. The second term can be decomposed by Cauchy-Schwartz inequality and non-amplifying of the filter functions as
\[
\left\| f^p_l - f'^p_l \right\| \leq \sum_{q=1}^{F_l-1} C_{\text{per}} \epsilon \| f^q_{l-1} \| + \sum_{q=1}^{F_l-1} \| f^q_{l-1} - f'^q_{l-1} \|,
\]
where \( C_{\text{per}} \) representing the constant in the stability bound of manifold filters. To solve this recursion, we need to compute the bound for \( \| f^q_p \| \). By normalized Lipschitz continuity of \( \sigma \) and the fact that \( \sigma(0) = 0 \), we can get
\[
\left\| f^q_p \right\| \leq \sum_{q=1}^{F_l-1} \| h^{pq}_l(\mathcal{L}) \| \left\| f^q_{l-1} \right\|
\leq \sum_{q=1}^{F_l-1} \| f^q_{l-1} \| \leq \sum_{q=1}^{F_l-1} \| f^q_q \|.
\]

Insert this conclusion back to solve the recursion, we can get
\[
\left\| f^p_l - f'^p_l \right\| \leq \| C_{\text{per}} \epsilon \left\| \prod_{l'=1}^{l-1} F_{l'} \right\| \left\| \sum_{q=1}^{F_l-1} \| f^q_q \| \right\|.
\]

Replace \( l \) with \( L \) we can obtain
\[
\left\| \Phi(H, \mathcal{L}, f) - \Phi(H, \mathcal{L}', f) \right\| \leq \sum_{q=1}^{F_L} \left( L C_{\text{per}} \epsilon \left\| \prod_{l'=1}^{L-1} F_{l'} \right\| \left\| \sum_{q=1}^{F_l} \| f^q_q \| \right\| \right).
\]

With \( F_0 = F_L = 1 \) and \( F_l = F \) for \( 1 \leq l \leq L - 1 \), then we have
\[
\left\| \Phi(H, \mathcal{L}, f) - \Phi(H, \mathcal{L}', f) \right\| \leq LF^{L-1} C_{\text{per}} \epsilon \| f \|,
\]
which concludes the proof.
Supplemental Materials

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 \( \mathcal{L}' = \mathcal{L} + A \) satisfy

\[
|\lambda_i - \lambda_i'| \leq \|A\|, \text{ for all } i = 1, 2, \ldots \tag{101}
\]

**Proof of Lemma 1**

The minimax principle asserts that

\[
\lambda_i(\mathcal{L}) = \max_{\text{codim } T = i-1} \min_{\mathcal{L} \leq u \in T, \|u\|=1} \langle \mathcal{L}u, u \rangle = \max_{\text{codim } T = i-1} \min_{\mathcal{L} \leq u \in T, \|u\|=1} \langle (\mathcal{L} + A)u, u \rangle = \max_{\text{codim } T = i-1} \min_{\mathcal{L} \leq u \in T, \|u\|=1} \langle \mathcal{L}u, u \rangle + \min_{\text{codim } T = i-1} \min_{\mathcal{L} \leq u \in T, \|u\|=1} \langle Au, u \rangle = \lambda_i(\mathcal{L}) + \lambda_1(A).	ag{102}
\]

Then for any \( 1 \leq i \), we have

\[
\lambda_i(\mathcal{L}') = \max_{\text{codim } T \leq i-1} \min_{\mathcal{L}' \leq u \in T, \|u\|=1} \langle (\mathcal{L} + A)u, u \rangle = \max_{\text{codim } T \leq i-1} \min_{\mathcal{L}' \leq u \in T, \|u\|=1} \langle (\mathcal{L} + A)u, u \rangle = \lambda_i(\mathcal{L}) + \lambda_1(A).	ag{103}
\]

Similarly, we can write \( \lambda_i(\mathcal{L}') \leq \lambda_i(\mathcal{L}) + \max_k \lambda_k(A) \). This leads to

\[
\lambda_1(A) + \lambda_1(A) = \lambda_i(\mathcal{L}) + \lambda_1(A).	ag{104}
\]

For the second term, we have

\[
|\langle \mathcal{L}u, u \rangle| \leq |\langle \mathcal{L}u, u \rangle| = \sum_i |\lambda_i(\mathcal{L})||\xi_i|^2 \tag{114}
\]

\[
\leq \max_n |\lambda_n(\mathcal{E})| \sum_i |\lambda_i(\mathcal{L})||\xi_i|^2 \tag{115}
\]

\[
|\langle \mathcal{L}u, u \rangle| \leq \varepsilon |\langle \mathcal{L}u, u \rangle| \tag{116}
\]

Therefore, we have

\[
\lambda_i(\mathcal{L} + \mathcal{E}) = \lambda_i(\mathcal{L}) + \varepsilon \max \min_{\text{codim } T \leq i-1} \langle \mathcal{L}u, u \rangle = \lambda_i(\mathcal{L}) + \varepsilon \lambda_i(\mathcal{L}), \tag{117}
\]

\[
\lambda_i(\mathcal{L} + \mathcal{E}) \geq \lambda_i(\mathcal{L}) - \varepsilon |\lambda_i(\mathcal{L})|, \tag{118}
\]

\[
\lambda_i(\mathcal{L}) - \varepsilon |\lambda_i(\mathcal{L})| \leq \lambda_i(\mathcal{L} + \mathcal{E}) \leq \lambda_i(\mathcal{L}) + \varepsilon |\lambda_i(\mathcal{L})|, \tag{119}
\]

which concludes the proof.

**F. Proof of Proposition 6**

We first import the existing results from [41] which indicates the spectral convergence of the constructed Laplacian operator based on the discretized manifold to the LB operator of the underlying manifold.

**Theorem 7 (Theorem 2.1 [41])** Let \( \lambda_i^n \) be the \( i \)-th eigenvalue of \( \mathcal{L}_n \) and \( \phi_i^n \) be the corresponding eigenfunction. Let \( \lambda_i \) and \( \phi_i \) be the corresponding eigenvalue and eigenfunction of \( \mathcal{L} \) respectively. Then there exists a sequence \( t_n \rightarrow 0 \), such that

\[
\lim_{n \rightarrow \infty} \lambda_i^n = \lambda_i, \quad \lim_{n \rightarrow \infty} \|\phi_i^n - \phi_i\| = 0, \tag{120}
\]

where the limits are taken in probability.

With the definitions of neural networks on discretized manifold and manifold, the output difference can be written as

\[
\mathcal{P}(\mathcal{H}, \mathcal{L}_n, \mathcal{P}, \mathcal{Q}_n(\mathcal{H}, \mathcal{L}_n, f)) = \sum_{q=1}^{F} \left| \sum_{l=1}^{l_{1}} \mathcal{P}_{q} - \sum_{l=1}^{l_{1}} \mathcal{Q}_{n} f_{l}^{q} \right|^2 \leq \left| \sum_{q=1}^{F} \left| \sum_{l=1}^{l_{1}} \mathcal{P}_{q} - \sum_{l=1}^{l_{1}} \mathcal{Q}_{n} f_{l}^{q} \right|^2 \right| \tag{121}
\]

By inserting the definitions, we have

\[
\mathcal{P}(\mathcal{H}, \mathcal{L}_n, \mathcal{P}, \mathcal{Q}_n(\mathcal{H}, \mathcal{L}_n, f)) = \left| \mathcal{F} \left( \sum_{q=1}^{F} \mathcal{P}_{q} \right) - \mathcal{F} \left( \sum_{q=1}^{F} \mathcal{Q}_{n} f_{l}^{q} \right) \right|^2 \tag{122}
\]

with \( x_0 = \mathcal{P}_1 f_1 \) as the input of the first layer. With a normalized Lipschitz nonlinearity, we have

\[
\left| \mathcal{P}_1 f_1 \right| \leq \left| \sum_{q=1}^{F} \mathcal{P}_{q} \right| - \mathcal{P}_1 \sum_{q=1}^{F} \mathcal{Q}_{n} f_{l}^{q} \tag{123}
\]

\[
\left| \mathcal{P}_1 f_1 \right| \leq \sum_{q=1}^{F} \left| \mathcal{Q}_{n} f_{l}^{q} \right| \tag{124}
\]
The difference can be further decomposed as
\[
\|h_i^{pq}(L_n)x_{i-1}^q - P_nh_i^{pq}(L)f_{i-1}^q\|
\leq \|h_i^{pq}(L_n)x_{i-1}^q - h_i^{pq}(L_n)P_nf_{i-1}^q
+ h_i^{pq}(L_n)P_nf_{i-1}^q - P_nh_i^{pq}(L)f_{i-1}^q\|
\leq \|h_i^{pq}(L_n)x_{i-1}^q - h_i^{pq}(L_n)P_nf_{i-1}^q\|
+ \|h_i^{pq}(L_n)P_nf_{i-1}^q - P_nh_i^{pq}(L)f_{i-1}^q\|
\]
(125)

The first term can be bounded as \(\|x_{i-1}^q - Pnf_{i-1}^q\|\) with the initial condition \(\|x_0 - Pnf_0\| = 0\). The second term can be denoted as \(D_i^0\). With the iteration employed, we can have
\[
\|\Phi(H, L_n, P_n f) - P_n\Phi(H, L, f)\| \leq \sum_{l=0}^{L} \prod_{l'=l}^{L} F_{l'} D_i^0.
\]
Therefore, we can focus on the difference term \(D_i^0\). We omit the feature and layer index to work on a general form, which leads to
\[
\|h(L_n)P_n f - P_nh(L)f\|
\leq \left\| \sum_{i=1}^{n} h(\lambda_i^n)\langle P_n f, \phi_i^n \rangle \phi_i^n
- \sum_{i=1}^{\infty} h(\lambda_i)\langle f, \phi_i \rangle P_n \phi_i \right\|
\leq \sum_{i=1}^{n} |h(\lambda_i^n)||\langle P_n f, \phi_i^n \rangle \phi_i^n
- \langle f, \phi_i \rangle P_n \phi_i | + \left\| P_n f \right\| \sum_{i=1}^{\infty} |h(\lambda_i) - h(\lambda_i^n)| + \left\| \sum_{i=n+1}^{\infty} h(\lambda_i)\langle f, \phi_i \rangle P_n \phi_i \right\|
\]
(127)

The first term in \(128\) can be bounded combined with the boundedness of frequency response as
\[
\sum_{i=1}^{n} \langle P_n f, \phi_i^n \rangle \|\phi_i^n - P_n \phi_i \| + \|\langle P_n f, \phi_i^n \rangle - \langle f, \phi_i \rangle \| \|\phi_i \|
\]
(129)

With the results in Theorem\([7]\) and \(\lim_{n \to \infty} P_n f = f\), the first term goes to 0 as \(n\) goes to infinity. The second term can be bounded by the Lipschitz continuity of the frequency response function, which leads to
\[
\left\| P_n f \right\| \sum_{i=1}^{n} |h(\lambda_i) - h(\lambda_i^n)| \leq C \left\| P_n f \right\| \sum_{i=1}^{n} |\lambda_i - \lambda_i^n|.
\]
(130)

As \(n \to \infty\), \(\lambda_i^n \to \lambda_i\). The last term also goes to 0 as \(n \to \infty\).
With \(D_i^0 \to 0\), this concludes the proof.