Stability of Neural Networks on Riemannian Manifolds

Zhiyang Wang, Luana Ruiz and Alejandro Ribeiro

Abstract—**Convolutional Neural Networks (CNNs) have been applied to data with underlying non-Euclidean structures and have achieved impressive successes. This brings the stability analysis of CNNs on non-Euclidean domains into notice because CNNs have been proved stable on Euclidean domains. This paper focuses on the stability of CNNs on Riemannian manifolds. By taking the Laplace-Beltrami operators into consideration, we construct an α-frequency difference threshold filter to help separate the spectrum of the operator with an infinite dimensionality. We further construct a manifold neural network architecture with these filters. We prove that both the manifold filters and neural networks are stable under absolute perturbations to the operators. The results also implicate a trade-off between the stability and discriminability of manifold neural networks. Finally we verify our conclusions with numerical experiments in a wireless adhoc network scenario.**

Index Terms—**Deep neural networks, Riemannian manifolds, stability analysis**

I. Introduction

Convolutional neural networks (CNNs) are machine learning architectures made up of layers where each layer composes a bank of convolutional filters with a pointwise nonlinearity. On problems where data is Euclidean, they have become a popular architecture due to their impressive performance in tasks ranging from speech recognition [1] to computer vision [2], which is largely attributed to the fact that CNNs are provably stable [3]. But in the physical world, their application is limited because many problems deal with data that is non-Euclidean. This is the case, for instance, of resource allocation in wireless ad-hoc communication networks [4], detection and recognition in social networks [5] and prediction of influenza epidemic outbreaks [6].

In recent years, a myriad of extensions of CNNs to non-Euclidean domains have been proposed to fill that gap [7]–[10]. These architectures have been able to reproduce the successes of CNNs on Euclidean domains to a large extent [11], [12], which naturally sparks the question of whether non-Euclidean CNNs are also stable. In this paper, we aim to answer this question by analyzing the properties of CNNs defined on the most general type of non-Euclidean domain — the manifold. By focusing on manifolds, our analysis has the benefit of being broad enough so that it can also be particularized to more specific non-Euclidean domains such as graphs, which can be seen as manifold discretizations.

Explicitly, we study the stability of manifold CNNs to absolute perturbations of the Laplace-Beltrami operator $\mathcal{L}$ asso-

ciated with the manifold (Definition 1). We start by analyzing the stability of the convolution operation, which is defined as a pointwise operation on the spectrum of $\mathcal{L}$. Given that absolute perturbations of $\mathcal{L}$ spawn absolute perturbations to all of its eigenvalues, designing stable manifold convolutions is challenging because the spectrum of the Laplace-Beltrami operator is infinite-dimensional. We address this challenge by introducing frequency difference threshold (FDT) filters (Definition 3), which separate the Laplacian spectrum into groups of eigenvalues that are less than some threshold α apart. We then show that these filters are stable to absolute perturbations of $\mathcal{L}$ (Theorem 1), and that this property is inherited by manifold CNNs (Theorem 2). The main implication of these results is that there is a trade-off between the stability and discriminability of manifold neural networks in the form of the frequency difference threshold of the FDT filters.

Related work includes a comprehensive study of the stability of graph neural networks (GNNs) in [13] and [14], which consider absolute and relative perturbations of the graph structure respectively, and the GNN stability analysis in [15], which focuses on perturbations of the graph spectrum. More in line with our paper, [16] studies stability of GNNs to perturbations of the underlying graph model, which is assumed to be a graphon. Unlike manifolds, however, graphons can only model dense graphs. More flexible models such as continuous graph models with tunable sparsity and generic topological spaces have been considered in [17] and [18] respectively, but these papers focus on the transferability and not on the stability of convolutional neural networks in these domains.

The rest of this paper is organized as follows. We start with a brief review of manifolds and Laplacian operators in Section II. We further introduce the framework of manifold convolutions and neural networks. In Section III, we introduce FDT filters and prove that the filters are stable under absolute perturbations of the Laplacian operator. We then extend this analysis to neural networks. Our results are verified numerically on a power allocation problem in wireless adhoc networks in Section IV, and conclusions are presented in Section V.

II. Preliminary definitions

In order to analyze the stability properties of neural networks on manifolds, we start by reviewing the notions of signals, Laplacian operators and convolutional neural networks in these domains.

A. Manifolds and manifold signals

A manifold is a topological space that is locally Euclidean around each point. More formally, a differentiable d-

dimensional manifold is a topological space where each point has a neighborhood that is homeomorphic to a d-dimensional Euclidean space, i.e., the tangent space. A Riemannian manifold, denoted $(\mathcal{M}, g)$, is a real and smooth manifold $\mathcal{M}$ equipped with a positive definite inner product g_x on the tangent space $T_x\mathcal{M}$ at each point x. The collection of tangent spaces of all points on $(\mathcal{M}, g)$ is denoted as $T\mathcal{M}$, and the collection of scalar functions and tangent vector functions on $(\mathcal{M}, g)$ are denoted as $L^2(\mathcal{M})$ and $L^2(T\mathcal{M})$ respectively. In this paper, we consider compact Riemannian manifolds $\mathcal{M}$.
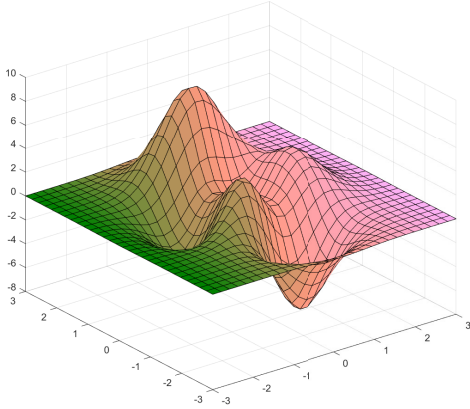

Figure 1: An example of manifold signal where the curve is the underlying manifold structure. Colors stand for values of the signal at the point.

Data supported on the manifold $\mathcal{M}$ is represented as manifold signals, which are defined as functions $f \in L^2(\mathcal{M})$. Given a manifold signal $f \in L^2(\mathcal{M}) : \mathcal{M} \to \mathbb{R}$, the Laplace-Beltrami operator $\mathcal{L}$ is defined as

$$\mathcal{L}f = -\text{div}(\nabla f), \tag{1}$$

where $\nabla : L^2(\mathcal{M}) \to L^2(T\mathcal{M})$ is an operator called *intrinsic gradient*. Compared with the classical notion of gradient, which indicates the direction of the fastest change of a function at a given point, the difference here is that the direction indicated by the intrinsic gradient is a tangent vector in $L^2(T\mathcal{M})$. The operator div is the *intrinsic divergence*, which is adjoint to the gradient operator. Akin to the Laplacian operator in Euclidean domains or the graph Laplacian in graph signal processing (GSP), the Laplace-Beltrami operator defines a notion of shift for signals on a manifold $\mathcal{M}$. In fact, $\mathcal{L}$ has a similar interpretation to that of the graph Laplacian as the difference between the local average function value around a point and the function value at the point itself [7].

The Laplacian operator $\mathcal{L}$ is a self-adjoint and positive-semidefinite operator. As such, it admits an eigendecomposition given by

$$\mathcal{L}f = \sum_{i=1}^{\infty} \lambda_i \langle f, \phi_i \rangle \phi_i. \tag{2}$$

where λ_i are real eigenvalues and ϕ_i are the corresponding eigenfunctions. Because the manifold $\mathcal{M}$ is compact, the spec-

trum $\{\lambda_i, \phi_i\}_{i \in \mathbb{N}^+}$ is additionally discrete. If the eigenvalues are ordered in increasing order as $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$, it can be shown that, for a manifold on d dimensions, λ_i grows as $i^{2/d}$ [19].

Similarly to how the eigenvalues of the graph Laplacian are interpreted as frequencies in GSP, we interpret the Laplacian-Beltrami operator eigenvalues λ_i as frequencies associated with oscillation modes ϕ_i. Moreover, since the eigenfunctions $\phi_1, \phi_2, \ldots$ form an orthonormal basis of $L^2(\mathcal{M})$, a change of basis operation for signals f can be defined, i.e., manifold signals $f \in L^2(\mathcal{M})$ can be represented on the manifold's eigenbasis as $f = \sum_{i=1}^{\infty} \langle f, \phi_i \rangle \phi_i$.

B. Manifold convolutions and manifold neural networks

Drawing a parallel with the spectral convolution operation on Euclidean domains, we leverage the eigendecomposition of the Laplacian (2) to define the convolution of manifold signals as a pointwise operation in the spectrum of $\mathcal{L}$. Explicitly, we define a manifold convolutional filter $\mathbf{h}(\mathcal{L})$ as

$$\mathbf{h}(\mathcal{L})f := \sum_{i=1}^{\infty} \sum_{k=0}^{K-1} h_k \lambda_i^k \langle f, \phi_i \rangle \phi_i \tag{3}$$

where $h_0, \ldots, h_{K-1}$ are the filter coefficients or taps.

Projecting $\mathbf{h}(\mathcal{L})f$ onto $\{\phi_i\}_{i \in \mathbb{N}^+}$, we see that the spectral response of the manifold convolution is is given by $h(\lambda) = \sum_{k=0}^{K-1} h_k \lambda^k$. This highlights the fact that the frequency response of a manifold convolution only depends on the coefficients h_k and the eigenvalues of the Laplace-Beltrami operator. It also implies that, if the underlying manifold changes — and thus $\mathcal{L}$ —, the behavior of the filter $\mathbf{h}$ can be replicated on the new manifold $\mathcal{M}'$ by evaluating $h(\lambda)$ at the eigenvalues of the new Laplacian $\mathcal{L}'$. Another important consequence of the spectral representation of $\mathbf{h}(\mathcal{L})$ being a polynomial is that, as $K \to \infty$, $\mathbf{h}(\mathcal{L})$ can be used to implement any smooth spectral response $h(\lambda)$ with convergent Taylor series around each $\lambda_0 \in \mathbb{R}$ as $K \to \infty$ [20].

From the definition of the convolution operation on $\mathcal{M}$ (3), convolutional neural networks (CNNs) are straightforward to define. A CNN consists of a cascade of L layers, each of which contains a bank of convolutional filters followed by a nonlinear activation function. Denoting the nonlinearity σ, the l-th layer of a L-layer CNN on the manifold $\mathcal{M}$ is given by:

$$f_l^p(x) = \sigma \left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q(x) \right) \tag{4}$$

for $l = 1, 2, \ldots, L$. Each of the filters $\mathbf{h}_l^{pq}(\mathcal{L})$ is as in (3) and maps the q-th feature from the $l-1$-th layer to the p-th feature of the l-th layer for $1 \leq q \leq F_{l-1}$ and $1 \leq p \leq F_l$. The output of this neural network is given by $y^p = f_L^p$ for $1 \leq p \leq F_L$ and the input features at the first layer, f_0^q, are the input data f^q for $1 \leq q \leq F_0$. For a more concise representation of this CNN, we can alternatively write it as the map $\mathbf{y} = \mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$, where $\mathbf{H}$ is a tensor gathering the learnable parameters $\mathbf{h}_l^{pq}$ at all layers of the CNN. We will refer to this map as a manifold convolutional neural network or manifold neural network (MNN) for short.

III. Stability of Manifold Neural Networks

In order to characterize the stability properties of MNNs, we first have to study the stability of their main component — the convolutional filter in (3). In particular, we analyze the stability of convolutional filters to absolute perturbations of the Laplace-Beltrami operator, which are specified in Definition 1.

Definition 1 (Absolute perturbations). Let $\mathcal{L}$ be the Laplace-Beltrami operator of a Riemannian manifold $\mathcal{M}$. An absolute perturbation of $\mathcal{L}$ is defined as

$$\mathcal{L}' = \mathcal{L} + \mathbf{A}, \tag{5}$$

where the absolute perturbation operator $\mathbf{A}$ is symmetric.

The absolute perturbation model introduced in Definition 1 only requires $\mathbf{A}$ to be symmetric. Thus, it is a rather generic perturbation model encompassing many different types of perturbations and, in particular, allowing to model a wide array of perturbations to the underlying manifold $\mathcal{M}$.

A. Frequency difference threshold (FDT) filters

Given the spectral decomposition of the Laplace-Beltrami operator (1), we can expect an absolute perturbation of $\mathcal{L}$ to spawn some sort of perturbation to the eigenvalues λ_i. Since the spectral convolution operation (3) depends on the evaluation of $h(\lambda)$ at each λ_i, its stability analysis will depend on the individual effects of the perturbation on each of these eigenvalues. A challenge in the case of manifolds is that the spectrum of $\mathcal{L}$ is infinite-dimensional, i.e., there is an infinite (albeit countable) number of eigenvalues λ_i. However, it is possible to show that these eigenvalues accumulate in certain parts of the spectrum. This is demonstrated in Proposition 1.

Proposition 1. Let $(\mathcal{M}, g)$ be a d-dimensional Riemannian manifold with Laplacian-Beltrami operator $\mathcal{L}$, and let λ_k denote the eigenvalues of $\mathcal{L}$. Let C_1 denote an arbitrary constant and let C_d be the volume of the d-dimensional unit ball. For any $\alpha > 0$, there exists N_1 given by

$$N_1 = \lceil (\alpha d/C_1)^{d/(2-d)}(C_d\text{Vol}(\mathcal{M}, g))^{2/(2-d)} \rceil \tag{6}$$

such that, for all $k > N_1$, it holds that

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

Proof. This is a direct consequence of Weyl's law [19]. $\square$

Proposition 1 is important because it suggests a strategy to mitigate the challenge posed by the infinite-dimensional spectrum of $\mathcal{L}$. Since eigenvalues accumulate in certain parts of this spectrum, for $\alpha > 0$ we can gather eigenvalues that are less than α apart in a *finite* number of groups. This α-separated spectrum, formalized in Definition 2, is achieved by the so-called frequency difference threshold (FDT) filters introduced in Definition 3.

Definition 2 (α-separated spectrum). The α-separated spectrum of a Laplace-Beltrami operator $\mathcal{L}$ is defined as the union of the set of α-separated eigenvalues $\mathcal{D}$,

$$\mathcal{D} = \{\lambda_i : i \geq 1, \ \lambda_i - \lambda_{i-1} > \alpha, \ \lambda_{i+1} - \lambda_i > \alpha\}$$
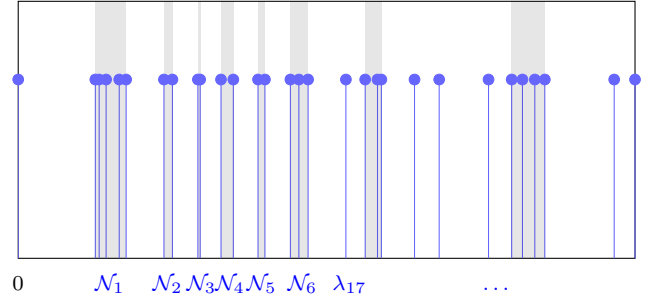

Figure 2: Eigenvalues of a Laplacian operator. Observe that the eigenvalues of the Laplacian operator tend to be grouped in certain parts of the spectrum.

and the set of α-close eigenvalues $\mathcal{N} = \bar{\mathcal{D}}$,

$$\mathcal{N} := \{\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_N\} \text{ s.t. } \min |\lambda_i - \lambda_j| > \alpha$$
$$\text{for all } \lambda_i \in \mathcal{N}_a, \lambda_j \in \mathcal{N}_b, a \neq b.$$

Definition 3 (α-FDT filter). On the manifold $\mathcal{M}$, an α-frequency difference threshold (α-FDT) filter is a filter whose frequency response $h(\lambda)$ satisfies

$$h(\lambda) = C_n, \quad \lambda \in [\min_{i \in \mathcal{N}_n} \lambda_i, \max_{j \in \mathcal{N}_n} \lambda_j]. \tag{7}$$

for all $n = 1, 2, \ldots, N$. This makes the convolution operation as:

$$\mathbf{h}(\mathcal{L}) = \sum_{i \in \mathcal{D}} h(\lambda_i)\langle f, \phi_i \rangle \phi_i + \sum_{n=1}^{N} C_n \langle f, E_n \rangle E_n, \tag{8}$$

where E_n is the eigenspace composed of $\{\phi_i\}_{i \in \mathcal{N}_n}$.

The eigenvalues λ_i in the set $\mathcal{D}$ are at least α apart from their preceding and succeeding eigenvalues λ_{i-1} and λ_{i+1}. Hence, according to Definition 3, an FDT filter with threshold α treats them as independent eigenvalues, like a conventional convolutional filter (9) would. The eigenvalues in the complement set $\mathcal{N} = \bar{\mathcal{D}}$ are divided in groups $\mathcal{N}_a$ where each group contains eigenvalues that are less than α apart from at least one their neighbors. As such, the distance between any two groups $\mathcal{N}_a, \mathcal{N}_b$, $a \neq b$, is larger than α. To achieve spectrum separation, the α-FDT filter imposes a constant frequency response $h(\lambda_i) = C_a$ for all $i \in \mathcal{N}_a$. In other words, it treats all the eigenvalues in a group $\mathcal{N}_a$ as the same. Note that, while the $h(\lambda)$ in Definition 3 is not a smooth function, we can still obtain a smooth approximation of an FDT filter in the form of (3).

B. FDT filter stability

Thanks to spectrum separation, α-FDT manifold filters can be shown to be stable under the absolute perturbations to the Laplace-Beltrami operator in Definition 1. We can state and prove this in Theorem 1 under Assumption 1.

Assumption 1. The filter function $h : \mathbb{R} \to \mathbb{R}$ is B-Lipschitz continuous and non-amplifying, i.e.,
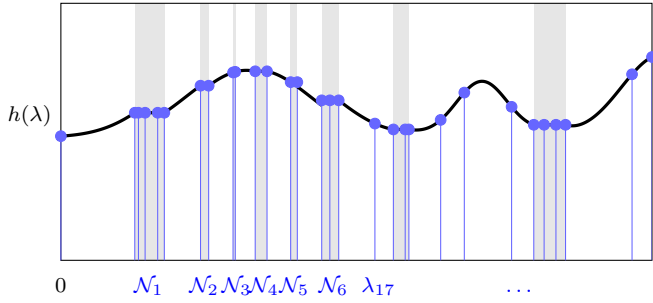
$$|h(a) - h(b)| \leq B|a - b|, \quad |h(a)| < 1. \tag{9}$$

Figure 3: An α-FDT filter that separates the spectrum of the Laplacian operator by grouping eigenvalues that are less than α apart in sets $\mathcal{N}_i$. Note that the frequency response is constant for all eigenvalues in a group $\mathcal{N}_i$.

Theorem 1 (FDT filter stability). Let $\mathcal{M}$ be a manifold with Laplace-Beltrami operator $\mathcal{L}$. Let f be a manifold signal and $\mathbf{h}(\mathcal{L})$ an α-FDT filter on $\mathcal{M}$ [cf. Definition 3]. Consider an absolute perturbation $\mathcal{L}' = \mathcal{L} + \mathbf{A}$ of the Laplace-Beltrami operator $\mathcal{L}$ [cf. Definition 1] where $\|\mathbf{A}\| = \epsilon \leq \alpha$. Then, under Assumption 1 it holds

$$\|\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathcal{L}')f\| \leq \frac{\pi(D+N)}{\alpha - \epsilon}\epsilon\|f\| + BD\epsilon\|f\| \quad (10)$$

where D and N are the cardinalities of sets $\mathcal{D}$ and $\mathcal{N}$.

Provided that Assumption 1 is satisfied, FDT filters are thus stable to absolute perturbations of the operator $\mathcal{L}$. Note that this assumption is rather reasonable, as there are no constraints on the value of the Lipschitz constant B. The stability bound depends on (i) the variability of the FDT filter as measured by the Lipschitz constant B and (ii) its frequency difference threshold, both directly through α, and indirectly through D and N. It is also split into two terms. The first one arises from the difference between the eigenfunctions of $\mathcal{L}$ and $\mathcal{L}'$, while the second is a result of the distance between their eigenvalues.

We observe that stability is improved if the Lipschitz constant B is small. However, this causes the filter to become less discriminative and give similar response to all spectral components. With a larger α, the FDT filter identifies fewer eigenvalues as being α-separated, which decreases D. While this reflects an increase in the number of groups N, a certain number of eigenvalues previously in $\mathcal{D}$ ends up being replaced by a fewer number of groups in $\mathcal{N}$, leading $D + N$ to also decrease. Thus, larger values of α improve stability, but this also happens at the cost of discriminability as filters with large α separate the spectrum more sparsely, i.e., they identify fewer eigenvalues as being α-separated. Interestingly, this stability-discriminability trade-off does not depend on the magnitude of the frequencies amplified by the filter (as is the case in, e.g., [14], [21]). Instead, it is associated with the filter's precision in telling neighboring frequencies apart wherever they mare in the spectrum.

C. Neural network stability

Manifold neural networks with banks of filters like the one in (7) inherit the stability properties of α-FDT filters. This is demonstrated in Theorem 2 under Assumption 2.

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

Theorem 2 (Neural network stability). Let $\mathcal{M}$ be a manifold with Laplace-Beltrami operator $\mathcal{L}$. Let f be a manifold signal and $\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$ an L-layer manifold neural network on $\mathcal{M}$ (4) with $F_0 = F_L = 1$ input and output features and $F_l = F, i = 1, 2, \ldots, L - 1$ features per layer, and where the filters $\mathbf{h}(\mathcal{L})$ are α-FDT filters [cf. Definition 3]. Consider an absolute perturbation $\mathcal{L}' = \mathcal{L} + \mathbf{A}$ of the Laplace-Beltrami operator $\mathcal{L}$ [cf. Definition 1] where $\|\mathbf{A}\| = \epsilon \leq \alpha$. Then, under Assumptions 1 and 2 it holds:

$$\begin{aligned}\|\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f) &- \mathbf{\Phi}(\mathbf{H}, \mathcal{L}', f)\| \\ &\leq LF^{L-1}\left(\frac{\pi(D+N)}{\alpha - \epsilon} + BD\right)\epsilon\|f\|.\end{aligned} \quad (11)$$

where D and N are the cardinalities of sets $\mathcal{D}$ and $\mathcal{N}$.

Deep neural networks on manifolds are thus also stable to absolute perturbations provided that the activation function is normalized Lipschitz. This assumption is satisfied by most common activation functions, such as the ReLU, the modulus function and the sigmoid. Here, the same general comments as in the case of Theorem 1 hold, with the difference that in Theorem 2 the stability bound also depends on the number of layers L and the number of features per layer F of the MNN.

As graphs can be seen as discretizations of manifolds, therefore we can use graph neural networks to realize the function of manifold neural networks. Combined with the transferability analysis from manifolds to graphs, we could state that the stability result that we have in Theorem 2 can be extended to graph neural networks.

IV. NUMERICAL EXPERIMENTS

We take a graph neural network model to approximate the manifold neural network. We verify our results on a graph neural network supported on a wireless adhoc network with $n = 50$ nodes within a range of $[-50m, 50m]^2$ where nodes are placed randomly. The channel states of all links can be represented by a matrix $\mathbf{H}(t)$ with each element $[\mathbf{H}(t)]_{ij} := h_{ij}(t)$ denotes the channel condition between node i and node j. Consider the large-scale pathloss gain and a random fast fading gain, this can be written as: $h_{ij} = \log(d_{ij}^{-2.2}h^f)$, where d_{ij} stands for the distance between node i and j, while $h^f \sim \text{Rayleigh}(2)$ is the random fading. We here study the power allocation problem among n nodes over an AWGN channel with interference, with $\mathbf{p}(\mathbf{H}) = [p_1(\mathbf{H}), p_2(\mathbf{H}), \ldots, p_n(\mathbf{H})]$ denoting the power allocated to each node under channel condition $\mathbf{H}$. the channel rate of node i is represented as r_i. The goal is to maximize the sum rate capacity under a total power budget P_{max}. This can
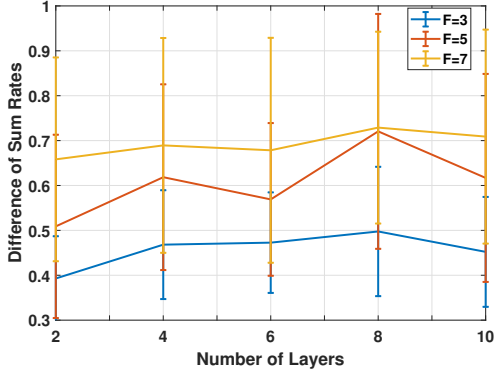
Figure 4: Sum-rate difference on the test between the original wirelss network setting and the perturbed one.

be numerically formulated as:

$$r^* = \max_{\mathbf{p}(\mathbf{H})} \sum_{i=1}^{n} r_i \qquad (12)$$

$$s.t. \quad r_i = \mathbb{E}\left[\log\left(1 + \frac{|h_{ii}|^2 p_i(\mathbf{H})}{1 + \sum_{j \neq i} |h_{ij}|^2 p_j(\mathbf{H})} \right) \right],$$

$$\mathbb{E}[\mathbf{1}^T \mathbf{p}] \leq P_{max}, \quad p_i(\mathbf{H}) \in \{0, p_0\}.$$

With the nodes and the links seen as the graph nodes and edges respectively, the channel matrix $\mathbf{H}$ can be seen as a graph shift operator, more specifically, an adjacency matrix. By formulating $\mathbf{H}$ into a Laplacian matrix, the problem can be solved with a graph neural network composed with our defined α-FDT filters with α set as 0.001. By setting $h_k = h_j$ for $|\lambda_k - \lambda_j| \leq \alpha$ in (3), we can get an approximation of α-FDT filters. After trained for 4000 iterations, the graph neural network can achieve the optimal power allocation. In physical world, the nodes are deployed in some specific spatial positions. The positions may change and this would cause perturbations to the underlying Laplacian matrix. To model this, we add a log-normal matrix to the original channel matrix $\mathbf{H}$. With the same trained graph neural network employed, we measure the stability by the difference of final sum-rate.

We can observe from Figure 4 that the difference of the final sum-rates increases with the number of layers and the number of filters per layer in the neural network. This verifies the conclusion that we have derived in Theorem 2.

V. CONCLUSIONS

In this paper, we have defined manifold convolutions and manifold neural networks. Considering the infinite dimensionality of Laplace-Beltrami operators. We import the definition of α frequency difference threshold filters to help separate the spectrum. By assigning a constant frequency response to the eigenvalues that are close enough, α-FDT filters can be proved to be stable to absolute perturbations to the Laplacian operators. We further prove that the manifold neural networks built with α-FDT filters are also stable under absolute perturbations. We conclude that there is a trade-off between the stability and

discriminability. We further verified our results numerically with a power allocation problem in wireless adhoc networks.

REFERENCES

[1] Wei Han, Zhengdong Zhang, Yu Zhang, Jiahui Yu, Chung-Cheng Chiu, James Qin, Anmol Gulati, Ruoming Pang, and Yonghui Wu, "Contextnet: Improving convolutional neural networks for automatic speech recognition with global context," *arXiv preprint arXiv:2005.03191*, 2020.

[2] Fredrik K Gustafsson, Martin Danelljan, and Thomas B Schon, "Evaluating scalable bayesian deep learning methods for robust computer vision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 318–319.

[3] Joan Bruna and Stéphane Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[4] Zhiyang Wang, Mark Eisen, and Alejandro Ribeiro, "Unsupervised learning for asynchronous resource allocation in ad-hoc wireless networks," *arXiv preprint arXiv:2011.02644*, 2020.

[5] Dat Nguyen, Kamela Ali Al Mannai, Shafiq Joty, Hassan Sajjad, Muhammad Imran, and Prasenjit Mitra, "Robust classification of crisis-related data on social networks using convolutional neural networks," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2017, vol. 11.

[6] Lijing Wang, Jiangzhuo Chen, and Madhav Marathe, "Defsi: Deep learning based epidemic forecasting with synthetic information," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 9607–9612.

[7] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[8] Fernando Gama, Antonio G Marques, Geert Leus, and Alejandro Ribeiro, "Convolutional neural network architectures for signals supported on graphs," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1034–1049, 2019.

[9] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

[10] Michaël Defferrard, Martino Milani, Frédérick Gusset, and Nathanaël Perraudin, "Deepsphere: a graph-based spherical cnn," *arXiv preprint arXiv:2012.15000*, 2020.

[11] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka, "How powerful are graph neural networks?," *arXiv preprint arXiv:1810.00826*, 2018.

[12] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, 2020.

[13] Fernando Gama, Joan Bruna, and Alejandro Ribeiro, "Stability of graph scattering transforms," *arXiv preprint arXiv:1906.04784*, 2019.

[14] Fernando Gama, Joan Bruna, and Alejandro Ribeiro, "Stability properties of graph neural networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5680–5695, 2020.

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

[16] Luana Ruiz, Zhiyang Wang, and Alejandro Ribeiro, "Graph and graphon neural network stability," *arXiv preprint arXiv:2010.12529*, 2020.

[17] Nicolas Keriven, Alberto Bietti, and Samuel Vaiter, "Convergence and stability of graph convolutional networks on large random graphs," *arXiv preprint arXiv:2006.01868*, 2020.

[18] Ron Levie, Michael M Bronstein, and Gitta Kutyniok, "Transferability of spectral graph convolutional neural networks," *arXiv preprint arXiv:1907.12972*, 2019.

[19] Wolfgang Arendt, Robin Nittka, Wolfgang Peter, and Frank Steiner, "Weyl's law: Spectral properties of the laplacian in mathematics and physics," *Mathematical analysis of evolution, information, and complexity*, pp. 1–71, 2009.

[20] Gordon K Smyth, "Polynomial approximation," *Wiley StatsRef: Statistics Reference Online*, 2014.

[21] Luana Ruiz, Luiz FO Chamon, and Alejandro Ribeiro, "Graphon neural networks and the transferability of graph neural networks," *arXiv preprint arXiv:2006.03548*, 2020.