Improved Stability Bounds of Graph Convolutional Neural Networks Under Graph Perturbations

Jun Zhang
School of Information Science and Technology
ShanghaiTech University
Shanghai, China
junzhang@shanghaitech.edu.cn

Ziping Zhao
School of Information Science and Technology
ShanghaiTech University
Shanghai, China
zipingzhao@shanghaitech.edu.cn

***Abstract*—Graph convolutional neural networks (GCNNs) have emerged as a powerful tool for processing signals or data supported on graphs. However, their effectiveness is compromised when perturbations exist in the graph structures. While previous research have studied the stability of GCNNs against such perturbations by analyzing the behavior of graph convolutional filters, we found a significant discrepancy exists between the theoretical stability bounds and simulation outcomes. In this paper, we propose a novel approach to characterize the stability of GCNNs more accurately. Unlike existing methods that treat graph convolutional filters as separate SISO systems, our approach views them as a compact MIMO system in each layer of a GCNN. This perspective yields an improved stability characterization, aligning more closely with empirical observations, and provides insights into designing GCNNs that are more resilient to graph perturbations. Numerical experiments on synthetic data verify our theoretical findings, and experiments on a movie recommendation problem demonstrate their utility in training stable GCNNs.**

***Index Terms*—Graph signal processing, graph convolution filters, graph convolutional neural networks, graph perturbations, stability.**

I. Introduction

Processing and learning from data in non-Euclidean domains have received increasing attention in recent years [1]–[3]. To model the underlying topology of data points in a non-Euclidean space, graphs have been adopted in many application scenarios, such as social networks [4], financial networks [5], wireless communication networks [6], molecular structures [7], etc. Among many graph-based methods, graph signal processing has emerged as an effective one. In graph signal processing, data is considered as signals supported on graphs and filters are defined to process graph signals [8]–[11], among which the graph convolutional filter (GCF), defined as a polynomial of some graph matrix, is a widely used one. In order to more accurately extract task-relevant representations, a multi-layered structure called graph convolutional neural network (GCNN) can be built, where each layer consists of a bank of GCFs and a nonlinear operation [12], [13].

In many applications, graph structures are prone to be perturbed by external factors such as estimation errors [14], [15], adversarial attacks [16], and variability of time [17]. When perturbed graphs are utilized to learn representations, one may obtain a degraded performance. This phenomenon necessitates studying the stability of graph-based models and devising stable architectures capable of accommodating variations of the underlying graph structure. The works [18], [19] studied the stability of a non-trainable graph neural networks called graph scattering transforms. For trainable architectures, [20] further investigated the impact of relative perturbations, demonstrating the stability and discriminative power of GCNNs at higher graph frequencies. Additionally, investigations conducted in [21]–[23] focused on the stability of GCNNs to structural perturbations with edge deletion or addition. Furthermore, the work in [24] studied the stability results to stochastic perturbations using random edge sampling. These stability analyses are of paramount importance in the development of resilient GCNN architectures and provide valuable theoretical insights into the fundamental principles underlying their functionality.

In particular, the stability of GCNNs has been shown to be directly effected by their integral Lipschitz constant of GCFs in the bank. However, we found that a substantial discrepancy exists between the existing theoretical stability bounds in [20] and the empirical outcomes. The reason for this discrepancy can be attributed to the fact that the current analysis treats the GCF banks in a GCNN as multiple individual single-input-single-output (SISO) GCFs and bounds the total perturbations of a GCNN by aggregating all the perturbations from individual GCFs. In this work, we aim to mitigate this gap between theoretical bounds and empirical observations via deriving a tighter stability bound for GCNNs. To address these limitation, we consider a bank of GCFs as a multiple-input-multiple-output (MIMO) GCF, and jointly analyze the stability property of the MIMO GCFs. Our result shows that both the integral Lipschitz constant and the frequency response bound of MIMO GCFs have influence on the stability of GCNNs.

The adoption of MIMO analysis in stability analysis is motivated by two key factors: the presence of cross-coupling effects and input correlations. In MIMO systems, inputs exhibit interdependencies, resulting in cross-coupling effects that extend beyond their corresponding output signals. This implies that each input signal influences not only its specific output but also other output signals. To accurately analyze errors and performance limitations, it is crucial to consider these

cross-coupling effects. MIMO analysis enables the capture and incorporation of mutual interactions between inputs and outputs, leading to more precise estimation of error bounds and system performance. Additionally, input correlations play a significant role in justifying the MIMO analysis. In a GCNN, the inputs may exhibit inherent correlations or interdependencies, meaning that variations or errors in one input can propagate and amplify across multiple outputs. Unlike SISO analysis, which assumes input independence, MIMO analysis explicitly accounts for input correlations and their impact on error propagation. By quantifying these correlations through the analysis of correlation matrices, MIMO analysis facilitates a more accurate assessment of the error amplification effects caused by input correlations. Considering input correlations in a comprehensive manner is essential for understanding system behavior and optimizing performance, providing insights that SISO analysis fails to capture.

Through experiments involving MIMO GCFs with random parameters, we demonstrate that our bound for the MIMO filter is tighter. Additionally, by regularizing both the Integral Lipschitz constant and the frequency response of the MIMO GCFs, we achieve better RSME performance and stability for a movie recommendation problem.

All the proofs of the results are available at: https://www.ncvxopt.com/pubs/ZhangZhao-GCNNStability.pdf.

II. Preliminaries on GCNN

A graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, which includes a set $\mathcal{V}$ with N nodes, a set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ and a weight function $\mathcal{W} : \mathcal{E} \to \mathbb{R}$. A graph signal is denoted as $\mathbf{x} = [x_1, \ldots, x_N]^{\mathsf{T}} \in \mathbb{R}^N$, which represents data associated with the nodes. The structure of a graph $\mathcal{G}$ is represented by a graph shift operator matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$, where $[\mathbf{S}]_{mn} \neq 0$ only if $(m, n) \in \mathcal{E}$ or $m = n$. Commonly used candidates of $\mathbf{S}$ are adjacency matrices, Laplacians, and their normalized counterparts [10]. In this paper, we focus on undirected graphs, in which case $\mathbf{S}$ is symmetric with eigendecomposition $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{T}}$. An efficient tool to process graph signal is the GCF. A GCF is defined as a polynomial on the graph shift operator $\mathbf{S}$. Specifically, a GCF of order K is defined as $\sum_{k=0}^{K} h_k \mathbf{S}^k$, where $\{h_k\}_{k=0}^{K}$ are filter coefficients. The output signal from a GCF with input $\mathbf{x}$ is represented as

$$\mathbf{y} = \sum_{k=0}^{K} h_k \mathbf{S}^k \mathbf{x}, \tag{1}$$

which is also a graph signal.

Equation (1) characterizes a single-input-single-output (SISO) system. Generally, we can consider a multiple-input-multiple-output (MIMO) system, where we have G input signals $\mathbf{x}^g \in \mathbb{R}^N$ for $g = 1, \ldots, G$, and F output signals $\mathbf{y}^f \in \mathbb{R}^N$ for $f = 1, \ldots, F$. The computation of the f-th output is based of a bank of GCFs, given by

$$\mathbf{y}^f = \sum_{g=1}^{G} \sum_{k=0}^{K} h_k^{fg} \mathbf{S}^k \mathbf{x}^g,$$

where h_k^{fg} is the k-th order coefficient associated with input $\mathbf{x}^g$ and output $\mathbf{y}^f$. Denote the input as $\mathbf{X} = \left[\mathbf{x}^1, \ldots, \mathbf{x}^G\right] \in \mathbb{R}^{N \times G}$, the output as $\mathbf{Y} = \left[\mathbf{y}^1, \ldots, \mathbf{y}^F\right] \in \mathbb{R}^{N \times F}$, and the filter coefficients as $\mathbf{H}_k \in \mathbb{R}^{G \times F}$ with $[\mathbf{H}_k]_{gf} = h_k^{fg}$, we can obtain a compact form of the MIMO GCF:

$$\mathbf{Y} = \sum_{k=0}^{K} \mathbf{S}^k \mathbf{X} \mathbf{H}_k. \tag{2}$$

If the input is a single graph signal, we have $G = 1$ and $\mathbf{X} = \mathbf{x} \in \mathbb{R}^{N \times 1}$. If the output is a single graph signal, we have $F = 1$ and $\mathbf{Y} = \mathbf{y} \in \mathbb{R}^{N \times 1}$. Thus, MIMO GCF can reduce to both two cases including SISO GCF.

GCNNs extend MIMO GCFs by using nonlinear functions that are applied independently to each component. Formally, we introduce a single variable function $\sigma : \mathbb{R} \to \mathbb{R}$ and extend it to a MIMO function $\sigma : \mathbb{R}^{n \times m} \to \mathbb{R}^{n \times m}$ by independent application to each component with $[\sigma(\mathbf{X})]_{ij} = \sigma\left([\mathbf{X}]_{ij}\right)$. Thus we use σ to denote both the scalar function and the pointwise function.

A GCNN is consists of a cascade of layers, each of which applies of an MIMO GCF followed by a nonlinear operation. Let F_ℓ denote the number of features for the ℓ-th layer. The input of layer ℓ is $\mathbf{X}_{\ell-1} \in \mathbb{R}^{N \times F_{\ell-1}}$, which is the output of layer $\ell - 1$. Then, the MIMO GCF's output $\mathbf{Y}_{\ell-1} \in \mathbb{R}^{N \times F_{\ell-1}}$ is followed by a pointwise nonlinear function σ to obtain the output

$$\mathbf{X}_\ell = \sigma(\mathbf{Y}_\ell). \tag{3}$$

Generally speaking, a GCNN defines a nonlinear map $\mathbf{\Phi}(\mathbf{S}, \mathbf{X}_0)$, which is the output of a cascade of layers applying to input signals $\mathbf{X} = \mathbf{X}_0$ supported on a graph $\mathbf{S}$. Each layer ℓ sequentially applies (2) and (3) with learnable parameters $\{\mathbf{H}_{\ell k}\}_k$.

III. An Improved Stability Analysis of GCNN

It is proved that the stability of GCFs to relative graph perturbations be directly effected by their integral Lipschitz constant. GCNNs inherit these properties since the central components of GCNNs are GCF banks [20]. Different from the work in [20], we consider a GCF bank as a MIMO GCF and achieve stability by regulating the integral Lipschitz constant of the MIMO filter.

A. Fundamentals

Consider a GCNN $\mathbf{\Phi}(\mathbf{S}, \mathbf{X}_0)$ with L layers, where the ℓ-th layer applies a MIMO GCF with coefficients $\{\mathbf{H}_{\ell k}\}_{k=0}^{K}$ and a nonlinear operation function $\boldsymbol{\sigma}(\cdot)$. Without loss of generality, we further assume that the nonlinear function is normalized Lipschitz so that $|\sigma(b) - \sigma(a)| \leq C_\sigma |b - a|$ with $C_\sigma = 1$. The stability of a GCNN is quantified by the difference in its outputs resulting from the application of a clean graph $\mathbf{S}$ and

a perturbed one $\hat{\mathbf{S}}$. Following [20], we consider a set of error matrices

$$\mathcal{RE}\left(\mathbf{S},\hat{\mathbf{S}}\right)=\{\mathbf{E}\mid\mathbf{P}^{\mathsf{T}}\hat{\mathbf{S}}\mathbf{P}=\mathbf{S}+\mathbf{E}\mathbf{S}+\mathbf{S}\mathbf{E},\mathbf{E}=\mathbf{E}^{\mathsf{T}}\},$$

where $\mathbf{P}$ is the permutation matrix. $\mathbf{E}$ reflects the relative changes of the graph structure, since $[\mathbf{E}\mathbf{S}+\mathbf{S}\mathbf{E}]_{mn}$ is proportional to the weighted sum of the degrees of nodes m and n scaled by the entries of $\mathbf{E}$. We assume there exists a weight matrix $\mathbf{E}\in\mathcal{RE}\left(\mathbf{S},\hat{\mathbf{S}}\right)$ that is both bounded by a small value and well-conditioned. Specifically, defining $d\left(\mathbf{S},\hat{\mathbf{S}}\right)=\min_{\mathbf{E}\in\mathcal{RE}\left(\mathbf{S},\hat{\mathbf{S}}\right)}\|\mathbf{E}\|_2$ as the difference between $\mathbf{S}$ and $\hat{\mathbf{S}}$, where $\|\cdot\|_2$ is the induced 2-norm. We can assume that there exists a weight matrix $\mathbf{E}\in\mathcal{RE}\left(\mathbf{S},\hat{\mathbf{S}}\right)$ which satisfies the following assumptions:

Assumption 1. $d\left(\mathbf{S},\hat{\mathbf{S}}\right)\leq\|\mathbf{E}\|_2\leq\epsilon$.

Assumption 2. *Eigenvalues* $\{\pi_n\}_{n=1}^N$ of $\mathbf{E}$ *are organized in order such that* $|\pi_1|\leq|\pi_2|\leq\cdots\leq|\pi_N|$, *satisfying* $|\pi_i/\pi_N-1|\leq\epsilon$.

In Assumption 1, the proportion of perturbation to the real graph is bounded by $\|\mathbf{E}\|_2$, while $|\pi_i/\pi_N-1|\leq\epsilon$ in Assumption 2 assumes that the difference between the proportion of perturbation to different graph components is bounded, so that it holds uniformly for all graphs independently of their number of nodes [20].

Then we study how a perturbation to $\mathbf{S}$ reflects on the output of a GCNN. Consider the permutation matrix $\mathbf{P}$, the difference of the output is given by:

$$\left\|\mathbf{P}^{\mathsf{T}}\mathbf{\Phi}\left(\hat{\mathbf{S}},\mathbf{P}\mathbf{X}_0\right)-\mathbf{\Phi}\left(\mathbf{S},\mathbf{X}_0\right)\right\|_{\mathsf{F}}. \tag{4}$$

Applying GCFs to the relabeled graph results in an output that corresponds to the relabeled output prior to any graph permutation. The following theorem asserts that GCNNs still hold the property.

Theorem 1 (Permutation Equivariance [20])**.** *Consider a permutation matrix* $\mathbf{P}\in\mathcal{P}$, *the permutation of the shift operator is* $\mathbf{P}^{\mathsf{T}}\mathbf{S}\mathbf{P}$ *and the permutation of the input data* $\mathbf{P}^{\mathsf{T}}\mathbf{X}$. *For a GCNN* $\mathbf{\Phi}(\cdot)$, *it holds that*

$$\mathbf{\Phi}\left(\mathbf{P}\mathbf{S}\mathbf{P}^{\mathsf{T}},\mathbf{P}\mathbf{X}_0\right)=\mathbf{P}\mathbf{\Phi}\left(\mathbf{S},\mathbf{X}_0\right). \tag{5}$$

Theorem 1 shows that GCNNs retain the permutation equivariance inherited from GCFs. It states that a node reordering results in a corresponding reordering of the GCNNs output, implying GCNNs are independent of node labeling. Thus we can assume that $\mathbf{P}$ is an identity matrix and $\hat{\mathbf{S}}=\mathbf{S}+\mathbf{E}\mathbf{S}+\mathbf{S}\mathbf{E}$ without loss of generality. We rewrite (4) as

$$\left\|\mathbf{\Phi}\left(\mathbf{S}+\mathbf{E}\mathbf{S}+\mathbf{S}\mathbf{E},\mathbf{X}_0\right)-\mathbf{\Phi}\left(\mathbf{S},\mathbf{X}_0\right)\right\|_{\mathsf{F}}. \tag{6}$$

B. Stability Analysis

In the case of SISO, a signal $\mathbf{x}$ can be transformed into the frequency domain as $\tilde{\mathbf{x}}=\mathbf{V}^{\mathsf{T}}\mathbf{x}$, where $\mathbf{V}$ is the eigenvector basis of $\mathbf{S}$. Then SISO GCF with parameters $\{h_k\}_{k=0}^K$ has the frequency response $h(\lambda)=\sum_{k=0}^K h_k\lambda^k$. This frequency response enables to analysis of the stability of SISO GCFs [20]. When considering the MIMO case, we first define the frequency response of MIMO GCFs as well.

A multi input graph signal $\mathbf{X}$ in the frequency domain can be denoted as

$$\tilde{\mathbf{X}}=\mathbf{V}^{\mathsf{T}}\mathbf{X}.$$

The signal associated to frequency λ_i is $\tilde{\mathbf{X}}_{i,:}^{\mathsf{T}}$, which is the transpose of the i-th row vector of $\tilde{\mathbf{X}}$. Then we have the output in the frequency domain:

$$\tilde{\mathbf{Y}}=\mathbf{V}^{\mathsf{T}}\sum_{k=0}^K\mathbf{S}^k\mathbf{X}\mathbf{H}_k=\sum_{k=0}^K\mathbf{\Lambda}^k\tilde{\mathbf{X}}\mathbf{H}_k.$$

The output signal in the frequency domain associated to frequency λ_i is

$$\tilde{\mathbf{Y}}_{i,:}^{\mathsf{T}}=\sum_{k=0}^K\lambda_i^k\mathbf{H}_k^{\mathsf{T}}\tilde{\mathbf{X}}_{i,:}^{\mathsf{T}},$$

where $\tilde{\mathbf{X}}_{i,:}$ and $\tilde{\mathbf{Y}}_{i,:}$ denote the i-th row of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, respectively. This motivates the definition of the frequency response of the MIMO GCF

$$\mathbf{H}(\lambda)=\sum_{k=0}^K\lambda^k\mathbf{H}_k^{\mathsf{T}}. \tag{7}$$

The i-th component $\tilde{\mathbf{Y}}_{i,:}^{\mathsf{T}}$ of the output signal in the frequency domain can be obtained by $\tilde{\mathbf{Y}}_{i,:}^{\mathsf{T}}=\mathbf{H}(\lambda)\tilde{\mathbf{X}}_{i,:}^{\mathsf{T}}$. Compared to the SISO frequency response $h(\lambda)$, $\mathbf{H}(\lambda)$ jointly processes multiple input and produces multiple output. If there is only single input and single output, $\mathbf{H}(\lambda)$ will reduce to $h(\lambda)$.

Similar to [20], we introduce two conditions on the graph filter, preventing the filter from amplifying energy and ensuring its frequency response does not change faster than linear.

Definition 2 (Bounded Filter)**.** Given a set of filter coefficients $\{\mathbf{H}_k\}_{k=0}^K$, its frequency response is given by (7). We say the filter is MIMO bounded by a constant $B>0$ if for all λ, $\|\mathbf{H}(\lambda)\|\leq B$.

With condition in Definition 2, it holds that

$$\begin{aligned}\left\|\sum_{k=0}^{\infty}\mathbf{S}^k\mathbf{X}\mathbf{H}_k\right\|_{\mathsf{F}}^2&=\left\|\sum_{k=0}^K\mathbf{\Lambda}^k\mathbf{V}^{\mathsf{T}}\mathbf{X}\mathbf{H}_k\right\|_{\mathsf{F}}^2\\&\leq\sum_{i=1}^N\left\|\sum_{k=0}^K\lambda_i^k\mathbf{H}_k^{\mathsf{T}}\tilde{\mathbf{X}}_{i,:}\right\|_2^2\leq B\|\mathbf{X}\|_{\mathsf{F}}^2,\end{aligned} \tag{8}$$

which is used to prevent the filter from amplifying the energy of signals.

Definition 3 (MIMO Integral Lipschitz GCF)**.** Given a set of filter coefficients $\{\mathbf{H}_k\}_{k=0}^K$ with frequency response given by (7). We say the filter is MIMO integral Lipschitz if there exists a constant $C > 0$ such that for all λ_1 and λ_2,

$$\|\mathbf{H}(\lambda_1) - \mathbf{H}(\lambda_2)\|_2 \leq C \frac{|\lambda_1 - \lambda_2|}{|\lambda_1 + \lambda_2|/2}, \tag{9}$$

Definition 3 states that $\mathbf{H}(\lambda)$ with constant C should exhibit slower changes as $|\lambda|$ increases. Specifically, for large values of λ the frequency response $\mathbf{H}(\lambda)$ must be flat and cannot discriminate nearby spectral features. But at values of $\lambda \approx 0$ the filters can be designed to produce output close to spectral features.

For relative perturbation models and GCNNs with MIMO integral Lipschitz filters the following stability result holds.

Theorem 4. *Let* $\mathbf{S}$ *and* $\hat{\mathbf{S}}$ *associated with two different graphs with the same node sets such that the error matrix* $\mathbf{E}$ *satisfies Assumption 1 and Assumption 2. Let the filter with parameters* $\{\mathbf{H}_k\}_{k=0}^K$ *be MIMO integral Lipschitz with a constant L, it holds that:*

$$\left\| \sum_{k=0}^{K} \hat{\mathbf{S}}^k \mathbf{X} \mathbf{H}_k - \sum_{k=0}^{K} \mathbf{S}^k \mathbf{X} \mathbf{H}_k \right\|_{\mathsf{F}} \leq 2C\epsilon \|\mathbf{X}\|_{\mathsf{F}} + \mathcal{O}\left(\epsilon^2\right). \tag{10}$$

Remark 5. There are some other perturbed assumptions of $\hat{\mathbf{S}}$ [20]. If the error matrix $\mathbf{E}$ does not satisfy the condition $|\pi_i/\pi_N - 1| \leq \epsilon$, the first term in (10) is given by $2C\left(1 + \delta\sqrt{N}\right)\epsilon \|\mathbf{X}\|_{\mathsf{F}}$, where δ is the eigenvector misalignment between $\mathbf{S}$ and $\mathbf{E}$, N is the number of nodes on the graph. If the perturbation model is an absolute perturbation model, error matrix is given by which is $\mathbf{E} = \mathbf{P}^{\mathsf{T}}\hat{\mathbf{S}}\mathbf{P} - \mathbf{S}$. In this case, the upper bound of the output difference is given by $C'\left(1 + \delta\sqrt{N}\right)\epsilon \|\mathbf{X}\|_{\mathsf{F}}$, where C' is a Lipschitz constant with $\|\mathbf{H}(\lambda_1) - \mathbf{H}(\lambda_2)\|_2 \leq C'|\lambda_1 - \lambda_2|$ which does not depend on the size of λ.

Theorem 6. *Let* $\mathbf{S}$ *and* $\hat{\mathbf{S}}$ *be two different graphs with the same node sets such that the error matrix* $\mathbf{E}$ *satisfies Assumption 1 and Assumption 2. Consider a GCNN* $\mathbf{\Phi}(\cdot,\cdot)$ *with* L *layers, where the* ℓ*-th layer applies a MIMO GCF with coefficients* $\{\mathbf{H}_{\ell k}\}_{k=0}^K$ *and a nonlinear operation function* $\boldsymbol{\sigma}$ *with* $|\boldsymbol{\sigma}(b) - \boldsymbol{\sigma}(a)| \leq |b - a|$*. Consider the general case, let the filter* $\{\mathbf{H}_{\ell k}\}_{k=0}^K$ *in* ℓ*-th layer with frequency response* $\mathbf{H}_\ell(\lambda) = \sum_{k=0}^K \lambda^k \mathbf{H}_{\ell k}^{\mathsf{T}}$ *be bounded by* B_ℓ *and integral Lipschitz with constant* C_ℓ*. That is, for all* λ_1 *and* λ_2

$$\|\mathbf{H}_\ell(\lambda_1) - \mathbf{H}_\ell(\lambda_2)\|_2 \leq C_\ell \frac{|\lambda_1 - \lambda_2|}{|\lambda_1 + \lambda_2|/2}$$
$$\|\mathbf{H}_\ell(\lambda_1)\|_2 \leq B_\ell.$$

Then, the difference between the output $\mathbf{X}_L$ *and the perturbed output* $\hat{\mathbf{X}}_L$ *is upper bounded by:*

$$\left\| \mathbf{\Phi}\left(\hat{\mathbf{S}}, \mathbf{X}_0\right) - \mathbf{\Phi}\left(\mathbf{S}, \mathbf{X}_0\right) \right\|_{\mathsf{F}} \leq 2 \sum_{\ell=1}^{L} \frac{C_\ell}{B_\ell} \prod_{i=1}^{L} B_i \epsilon \|\mathbf{X}_0\|_{\mathsf{F}} + \mathcal{O}\left(\epsilon^2\right).$$

Proof: The proof is given in the Supplementary Materials. ∎

The above theoretical result coincides with the results in [20] for GCNNs equipped with a SISO filter in each layer but differs for GCNNs with MIMO filters. C_ℓ and B_ℓ are only determined by the parameters of MIMO GCFs. Although reducing the hidden features is likely to get smaller C_ℓ and B_ℓ, it also reduces the representational capacity. To address this from a practical standpoint, we can add a penalty of C_ℓ and B_ℓ to the loss function when training the GCNNs.

IV. Numerical Simulations

A. MIMO GCF Stability

To evaluate the correctness and tightness of our bound, we compare the obtained stability bounds with the existing ones [20] through synthetic data. The results are given in Fig. 1a. We randomly generate 100 graphs with 10 nodes and their adjacency matrices $\mathbf{S}$. Both the input graph signals and the GCF coefficients are sampled from a standard Gaussian distribution. The error matrices $\mathbf{E}$ are chosen to be diagonal matrices with elements drawn uniformly from the interval $[\epsilon(1-\epsilon), \epsilon]$. Then we consider the output changes of GCFs under these perturbations. In Fig. 1a, we depict the output differences (averaged over 100 graphs) of both SISO GCFs (in red) and MIMO GCFs (in blue). We also compute the theoretical bounds of the output differences given by [20] (denoted as benchmark) and the result in Theorem 4 (denoted as proposed), where we only plot the first term. Fig. 1a shows that the stability bound in [20] is tight for the SISO case but differs largely for the MIMO case. By contrast, our derived bound for the MIMO case is much tighter. Besides, Fig. 1b shows that the bound in [20] grows superlinearly as the feature number increases, while the empirical output difference and our bound both increase linearly.

Both Fig. 1a and Fig. 1b show that our analysis of the stability of MIMO GCFs gives a tighter bound compared to the original bound. The reason why there is a gap for the MIMO case is due to the presence of cross-coupling effects and input correlations, which can be handled by extending the stability analysis to the MIMO case.

B. GCNN Stability

We also illustrate that our GCNN stability analysis can help to train a stable GCNN. This target is demonstrated based on using GCNN for movie recommendation on the MovieLens-100k dataset [25], which contains 100,000 ratings given by 943 users to 1,582 movies. One movie is modeled as a node in the graph, and the edge weights are computed as the Pearson correlation coefficient [26]. All ratings by one user are treated as a graph signal. We consider three different GCNN architectures mapping the known ratings to a target unknown one. All these GCNNs have two layers, where $F_0 = 1$ and $F_1 = F_2 = 32$, and each layer contains a MIMO GCF of order $K = 4$, followed by a ReLU activation function. Following the two-layer GCNN, there is a linear readout layer. The first

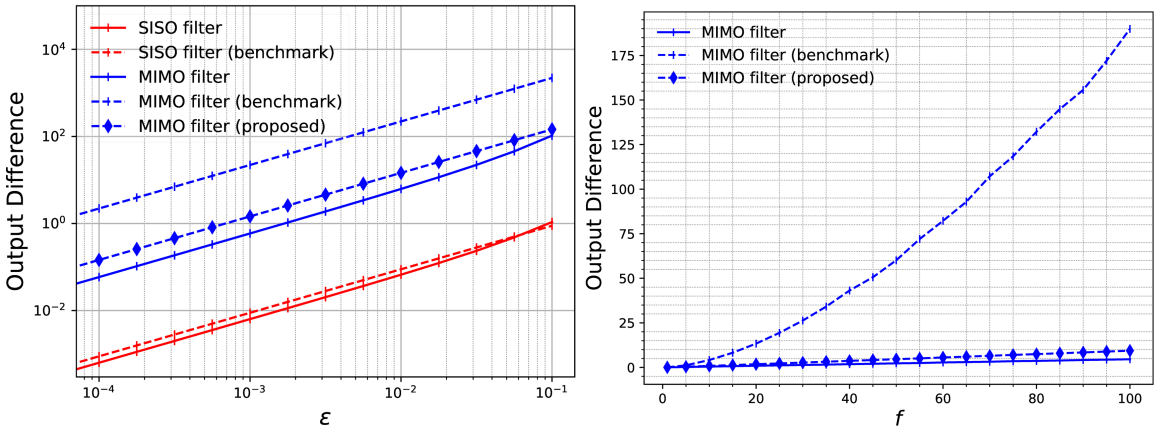

Fig. 1: (a) GCNN output difference by varying ϵ; (b) GCNN output difference by varying feature dimension.

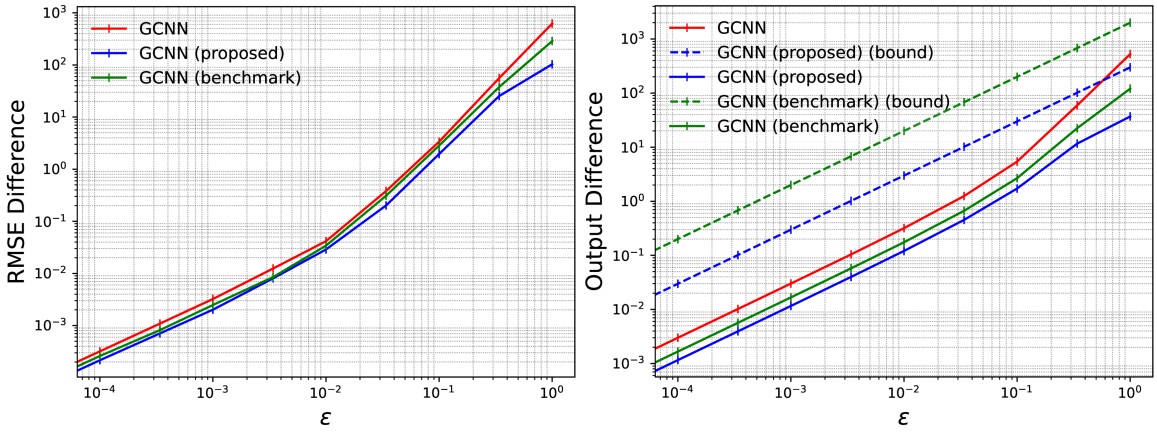

Fig. 2: (a) Change in RMSE by varying the relative error ϵ; (b) Change in the output of the GCNN by varying the relative error ϵ.

architecture, labeled as 'GCNN', learns from the entire space of GCFs. The second one, labeled as 'GCNN (benchmark)', learns GCFs based on [20], which are bounded and SISO integral Lipschitz. The third architecture, denoted as 'GCNN (proposed)', learns GCFs based on the proposed stability analysis, which are bounded and MIMO integral Lipschitz. We construct the error matrix $\mathbf{E}$ similarly to the synthetic experiments. In Fig. 2, we plot the GCNN output difference and the root mean square error (RMSE) of ratings by varying the perturbation size ϵ from 10^{-4} to 1. We also compute the theoretical bounds of the output difference given by [20] (denoted as GCNN (benchmark) (bound)) and Theorem 6 (denoted as GCNN (proposed) (bound)). Our findings indicate that for small ϵ, there is minimal output change across all three architectures. However, as ϵ increases, the GCNN with MIMO integral Lipschitz filters exhibits smaller changes compared to both the GCNN and GCNN (benchmark) architectures. Both Fig. 2a and Fig. 2b show that the GCNN (proposed) is more stable than the GCNN and GCNN (benchmark). Also, our bound to the change of GCNN output is tighter.

V. Conclusion

In this paper, we have introduced an improved stability characterization for GCNNs by regarding the linear feature transformation step in each layer of GCNN as a MIMO GCF. By considering the collective behavior of the filters in the bank, our approach enhances the stability of GCNNs. Experimental results on the graph convolutional filter stability and on a movie recommendation problem have validated the effectiveness of our method.

Our findings contribute to the field of graph signal processing and provide insights for designing more robust and stable GCNN architectures. Future research can explore further applications and extensions of the MIMO frameworks in other domains like convolutional neural networks for signals supported on multi-graph. Overall, our work advances the understanding and development of GCNNs, enabling their utilization in various applications requiring stable and reliable graph-based learning models.

REFERENCES

[1] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph Signal Processing: Overview, Challenges, and Applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[2] G. Leus, A. G. Marques, J. M. Moura, A. Ortega, and D. I. Shuman, "Graph Signal Processing: History, Development, Impact, and Outlook," *IEEE Signal Processing Magazine*, vol. 40, no. 4, pp. 49–60, Jun. 2023.

[3] X. Dong, D. Thanou, L. Toni, M. Bronstein, and P. Frossard, "Graph Signal Processing for Machine Learning: A Review and New Perspectives," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 117–127, Nov. 2020.

[4] C. Chamley, A. Scaglione, and L. Li, "Models for the Diffusion of Beliefs in Social Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 16–29, 2013.

[5] L. Di Persio, O. Honchar *et al.*, "Artificial Neural Networks Architectures for Stock Price Prediction: Comparisons and Applications," *International Journal of Circuits, Systems and Signal Processing*, vol. 10, pp. 403–413, 2016.

[6] M. Eisen and A. Ribeiro, "Optimal Wireless Resource Allocation With Random Edge Graph Neural Networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2977–2991, 2020.

[7] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, "The Rise of Deep Learning in Drug Discovery," *Drug Discovery Today*, vol. 23, no. 6, pp. 1241–1250, Jun. 2018.

[8] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs: Graph Filters," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6163–6166.

[9] A. Sandryhaila and J. M. Moura, "Discrete Signal Processing on Graphs: Graph Fourier Transform," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6167–6170.

[10] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[11] M. Cheung, J. Shi, O. Wright, L. Y. Jiang, X. Liu, and J. M. F. Moura, "Graph Signal Processing and Deep Learning: Convolution, Pooling, and Topology," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 139–149, Nov. 2020.

[12] F. Gama, E. Isufi, G. Leus, and A. Ribeiro, "Graphs, Convolutions, and Neural Networks: From Graph Filters to Graph Neural Networks," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 128–138, Nov. 2020.

[13] 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, Feb. 2019.

[14] P. V. Marsden, "Network Data and Measurement," *Annual Review of Sociology*, vol. 16, no. 1, pp. 435–463, 1990.

[15] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network Topology Inference from Spectral Templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[16] D. Zügner and S. Günnemann, "Certifiable Robustness and Robust Training for Graph Convolutional Networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019.

[17] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Filtering Random Graph Processes Over Random Time-Varying Graphs," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4406–4421, Aug. 2017.

[18] F. Gama, A. Ribeiro, and J. Bruna, "Diffusion scattering transforms on graphs," in *International Conference on Learning Representations*, 2018.

[19] F. Gama, J. Bruna, and A. Ribeiro, "Stability of Graph Scattering Transforms," *Advances in Neural Information Processing Systems*, 2019.

[20] ——, "Stability Properties of Graph Neural Networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5680–5695, 2020.

[21] H. Kenlay, D. Thanou, and X. Dong, "On the Stability of Polynomial Spectral Graph Filters," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing*. Barcelona, Spain: IEEE, May 2020, pp. 5350–5354.

[22] ——, "On The Stability of Graph Convolutional Neural Networks Under Edge Rewiring," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, Jun. 2021, pp. 8513–8517.

[23] X. Wang, E. Ollila, and S. A. Vorobyov, "Graph Neural Network Sensitivity under Probabilistic Error Model," in *2022 30th European Signal Processing Conference*, 2022, pp. 2146–2150.

[24] Z. Gao, E. Isufi, and A. Ribeiro, "Stability of Graph Convolutional Neural Networks to Stochastic Perturbations," *Signal Processing*, vol. 188, p. 108216, 2021.

[25] F. M. Harper and J. A. Konstan, "The Movielens Datasets: History and Context," *Acm Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, pp. 1–19, 2016.

[26] W. Huang, A. G. Marques, and A. R. Ribeiro, "Rating Prediction via Graph Signal Processing," *IEEE Transactions on Signal Processing*, vol. 66, no. 19, pp. 5066–5081, Oct. 2018.

VI. Proof Lemma 4

Proof: The proof is analogous to that for SISO case, following from the fact that graph filters are permutation equivariant, we can assume, without loss of generality, that $\mathbf{P} = \mathbf{I}$. From the first order expansion of $(\mathbf{S} + \mathbf{ES} + \mathbf{SE})^k$, we get

$$\begin{aligned}(\mathbf{S} + \mathbf{ES} + \mathbf{SE})^k &= \mathbf{S}^k \\ &+ \sum_{r=0}^{k-1} \left(\mathbf{S}^r \mathbf{E} \mathbf{S}^{k-r} + \mathbf{S}^{r+1} \mathbf{E} \mathbf{S}^{k-r-1}\right) + \mathbf{C},\end{aligned} \tag{11}$$

with $\mathbf{C}$ such that $\|\mathbf{C}\|_{\mathsf{F}} = \mathcal{O}(\epsilon^2)$. Using this first-order approximation in (10), we consider the difference in the effects of the filter on an arbitrary graph signal $\mathbf{X}$

$$\begin{aligned}&\sum_{k=0}^{K} \hat{\mathbf{S}}^k \mathbf{X} \mathbf{H}_k - \sum_{k=0}^{K} \mathbf{S}^k \mathbf{X} \mathbf{H}_k \\ &\quad = \sum_{k=0}^{K} \sum_{r=0}^{k-1} \left(\mathbf{S}^r \mathbf{E} \mathbf{S}^{k-r} + \mathbf{S}^{r+1} \mathbf{E} \mathbf{S}^{k-r-1}\right) \mathbf{X} \mathbf{H}_k + \mathbf{D}.\end{aligned} \tag{12}$$

With $\mathbf{D}$ such that $\|\mathbf{D}\|_{\mathsf{F}} / \|\mathbf{X}\|_{\mathsf{F}} = \mathcal{O}(\epsilon^2)$. Next, with the eigndecomposition $\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{\mathsf{T}}$, graph signal $\mathbf{X}$ has a GFT given by $\tilde{\mathbf{X}} = \mathbf{V}^{\mathsf{T}} \mathbf{X}$. Rewrite the first term in (12) as:

$$\begin{aligned}&\sum_{k=0}^{K} \sum_{r=0}^{k-1} \left(\mathbf{S}^r \mathbf{E} \mathbf{S}^{k-r} + \mathbf{S}^{r+1} \mathbf{E} \mathbf{S}^{k-r-1}\right) \mathbf{X} \mathbf{H}_k \\ &\quad = \sum_{k=0}^{K} \sum_{r=0}^{k-1} \left(\mathbf{S}^r \mathbf{E} \mathbf{V} \mathbf{\Lambda}^{k-r} + \mathbf{S}^{r+1} \mathbf{E} \mathbf{V} \mathbf{\Lambda}^{k-r-1}\right) \tilde{\mathbf{X}} \mathbf{H}_k\end{aligned}$$

We focus on the product $\mathbf{EV}$, let $\mathbf{E} = \mathbf{U} \mathbf{\Pi} \mathbf{U}^{\mathsf{T}}$, $\mathbf{u}_i$ is the eigenvector associated with i-th eigenvalue π_i. Then,

$$\mathbf{EV} = \sum_{i=1}^{N} \pi_i \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} \mathbf{V} = \pi_N \sum_{i=1}^{N} \frac{\pi_i}{\pi_N} \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} \mathbf{V}.$$

Then, recall that $|\pi_i / \pi_N - 1| \leq \epsilon$, we can write $\pi_i / \pi_N = \delta_i + 1$ with $\delta_i \leq \epsilon$, which yields

$$\mathbf{EV} = \pi_N \mathbf{V} + \pi_N \mathbf{W}, \qquad \mathbf{W} = \sum_{i=1}^{N} \delta_i \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} \mathbf{V}. \tag{13}$$

Note that

$$\|\mathbf{W}\|_2 \leq \left\| \sum_{i=1}^{N} \delta_i \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} \right\|_2 \|\mathbf{V}\|_2 \leq \epsilon.$$

Now, using (12) yields three terms

$$\sum_{k=0}^{K} \sum_{r=0}^{k-1} \left(\mathbf{S}^r \mathbf{E} \mathbf{V} \mathbf{\Lambda}^{k-r} + \mathbf{S}^{r+1} \mathbf{E} \mathbf{V} \mathbf{\Lambda}^{k-r-1}\right) \tilde{\mathbf{X}} \mathbf{H}_k \tag{14}$$

$$= 2\pi_N \mathbf{V} \sum_{k=0}^{K} k \mathbf{\Lambda}^k \tilde{\mathbf{X}} \mathbf{H}_k \tag{15}$$

$$+ \pi_N \sum_{k=0}^{K} \sum_{r=0}^{k-1} \mathbf{S}^r \mathbf{W} \mathbf{\Lambda}^{k-r} \tilde{\mathbf{X}} \mathbf{H}_k \tag{16}$$

$$+ \pi_N \sum_{k=0}^{K} \sum_{r=1}^{k} \mathbf{S}^r \mathbf{W} \mathbf{\Lambda}^{k-r} \tilde{\mathbf{X}} \mathbf{H}_k. \tag{17}$$

Consider the Frobenius norm of (15), denote $\tilde{\mathbf{X}}_{i,:}$ as the i-th row vector of $\tilde{\mathbf{X}}$, we obtain

$$\begin{aligned}\left\| 2\pi_N \mathbf{V} \sum_{k=0}^{K} k \mathbf{\Lambda}^k \tilde{\mathbf{X}} \mathbf{H}_k \right\|_{\mathsf{F}}^2 &= \left\| 2\pi_N \sum_{k=0}^{K} k \mathbf{\Lambda}^k \tilde{\mathbf{X}} \mathbf{H}_k \right\|_{\mathsf{F}}^2 \\ &= 4\pi_N^2 \sum_{i=1}^{N} \left\| \sum_{k=0}^{K} k \lambda_i^k \mathbf{H}_k^{\mathsf{T}} \tilde{\mathbf{X}}_{i,:}^{\mathsf{T}} \right\|_2^2 \\ &\leq 4\pi_N^2 \sum_{i=1}^{N} \left\| \sum_{k=0}^{K} k \lambda_i^k \mathbf{H}_k^{\mathsf{T}} \right\|_2^2 \left\| \tilde{\mathbf{X}}_{i,:} \right\|_2^2,\end{aligned}$$

where $d\mathbf{H}(\lambda_i) = \sum_{k=0}^{K} k \lambda_i^{k-1} \mathbf{H}_k^{\mathsf{T}} d\lambda_i$ is the derivative of $\mathbf{H}(\lambda)$ evaluated at $\lambda = \lambda_i$. From hypothesis (9), we have that

$$\begin{aligned}&\left\| \sum_{k=0}^{K} k \lambda^k \mathbf{H}_k^{\mathsf{T}} \right\|_2 \\ &\quad = \lim_{\lambda_1, \lambda_2 \to \lambda} \frac{|\lambda_1 + \lambda_2| \, \|\mathbf{H}(\lambda_2) - \mathbf{H}(\lambda_1)\|}{2 |\lambda_2 - \lambda_1|} \leq C.\end{aligned}$$

Therefore,

$$\begin{aligned}&\sum_{i=1}^{N} \left\| \sum_{k=0}^{K} k \lambda_i^k \mathbf{H}_k^{\mathsf{T}} \right\|_2^2 \left\| \tilde{\mathbf{X}}_{i,:} \right\|_2^2 \\ &\quad \leq C^2 \sum_{i=1}^{N} \left\| \tilde{\mathbf{X}}_{i,:} \right\|_2^2 = C^2 \left\| \tilde{\mathbf{X}} \right\|_{\mathsf{F}}^2 = C^2 \|\mathbf{X}\|_{\mathsf{F}}^2.\end{aligned} \tag{18}$$

This given the bound of (15)

$$\left\| 2\pi_N \mathbf{V} \sum_{k=0}^{K} k \mathbf{\Lambda}^k \tilde{\mathbf{X}} \mathbf{H}_k \right\|_{\mathsf{F}} \leq 2C\epsilon \|\mathbf{X}\|_{\mathsf{F}}. \tag{19}$$

In the case of (16) and (17), these two terms are similar. Denote $\mathrm{diag}\left([0, \cdots, \lambda_i, \cdots, 0]\right)$ as Λ_i, we rewrite (16) as

$$\begin{aligned}&\pi_N \sum_{k=0}^{K} \sum_{r=1}^{k} \mathbf{S}^r \mathbf{W} \mathbf{\Lambda}^{k-r} \tilde{\mathbf{X}} \mathbf{H}_k \\ &\quad = \pi_N \mathbf{V} \sum_{i=1}^{N} \sum_{k=0}^{K} \sum_{r=0}^{k-1} \Lambda_i^r \mathbf{V}^{\mathsf{T}} \mathbf{W} \mathbf{\Lambda}^{k-r} \tilde{\mathbf{X}} \mathbf{H}_k.\end{aligned} \tag{20}$$

We observe that only the i-th row vector of $\sum_{k=0}^{K}\sum_{r=0}^{k-1}\Lambda_i^r\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{\Lambda}^{k-r}\tilde{\mathbf{X}}\mathbf{H}_k$ is nonzero, then we compute the bound of its Frobenius norm,

$$\begin{aligned}&\left\|\pi_N\mathbf{V}\sum_{i=1}^{N}\sum_{k=0}^{K}\sum_{r=0}^{k-1}\Lambda_i^r\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{\Lambda}^{k-r}\tilde{\mathbf{X}}\mathbf{H}_k\right\|_\mathsf{F}^2\\&\quad=\epsilon^2\sum_{i=1}^{N}\left\|\sum_{k=0}^{K}\sum_{r=0}^{k-1}\Lambda_i^r\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{\Lambda}^{k-r}\tilde{\mathbf{X}}\mathbf{H}_k\right\|_2^2.\end{aligned}\tag{21}$$

It is immediate that the i-th row vector of $\Lambda_i^r\mathbf{V}^\mathsf{T}$ is $\lambda_i^r\mathbf{v}_i^\mathsf{T}$, while the others are $\mathbf{0}^\mathsf{T}$. Then

$$\left\|\sum_{k=0}^{K}\sum_{r=0}^{k-1}\Lambda_i^r\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{\Lambda}^{k-r}\tilde{\mathbf{X}}\mathbf{H}_k\right\|_2^2\tag{22}$$

$$=\left\|\sum_{k=0}^{K}\mathbf{H}_k^\mathsf{T}\tilde{\mathbf{X}}^\mathsf{T}\sum_{r=0}^{k-1}\lambda_i^r\mathbf{\Lambda}^{k-r}\mathbf{W}^\mathsf{T}\mathbf{v}_i\right\|_2^2\tag{23}$$

$$\leq\epsilon^2\left\|\sum_{k=0}^{K}\mathbf{H}_k^\mathsf{T}\tilde{\mathbf{X}}^\mathsf{T}\sum_{r=0}^{k-1}\lambda_i^r\mathbf{\Lambda}^{k-r}\right\|_2^2.\tag{24}$$

Denote $\sum_{r=0}^{k-1}\lambda_i^r\mathbf{\Lambda}^{k-r}$ as $\mathbf{G}_{k,i}$, the j-th diagonal element of $\mathbf{G}_{k,i}$ is $[\mathbf{G}_{k,i}]_{jj}$, rewrite (24)

$$\begin{aligned}&\left\|\sum_{k=0}^{K}\mathbf{H}_k^\mathsf{T}\tilde{\mathbf{X}}^\mathsf{T}\mathbf{G}_{k,i}\right\|_\mathsf{F}^2\\&=\sum_{j=1}^{N}\left\|\sum_{k=0}^{K}[\mathbf{G}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}\tilde{\mathbf{X}}_{j,:}\right\|_2^2\\&\quad=\sum_{j=1}^{N}\left\|\sum_{k=0}^{K}[\mathbf{G}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}\right\|_2^2\left\|\tilde{\mathbf{X}}_{j,:}\right\|_2^2.\end{aligned}$$

Similarly denote $\sum_{r=1}^{k}\lambda_i^r\mathbf{\Lambda}^{k-r}$ as $\mathbf{Q}_{k,i}$, the summation of (16) and (17) is upper bounded by

$$\sum_{j=1}^{N}\left\|\sum_{k=0}^{K}[\mathbf{G}_{k,i}+\mathbf{Q}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}\right\|_2^2\left\|\tilde{\mathbf{X}}_{j,:}\right\|_2^2.\tag{25}$$

The key step on the proof of the term (16) is the bound $\sum_{k=0}^{K}[\mathbf{G}_{k,i}+\mathbf{Q}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}$. Consider the diagonal entries of $\mathbf{G}_{k,i}+\mathbf{U}_{k,i}$,

$$[\mathbf{G}_{k,i}+\mathbf{Q}_{k,i}]_{jj}=\sum_{r=0}^{k-1}\left[\lambda_i^r\lambda_j^{k-r}+\lambda_i^{r+1}\lambda_j^{k-r-1}\right].$$

For the case of $j=i$,

$$\begin{aligned}\sum_{k=0}^{K}[\mathbf{G}_{k,i}+\mathbf{Q}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}&=\sum_{k=0}^{K}\sum_{r=0}^{k-1}\left[\lambda_i^r\lambda_j^{k-r}+\lambda_i^{r+1}\lambda_j^{k-r-1}\right]\mathbf{H}_k^\mathsf{T}\\&=2\sum_{k=0}^{K}k\lambda_i^k\mathbf{H}_k^\mathsf{T}=\frac{2\lambda_i d\mathbf{H}(\lambda_i)}{d\lambda_i}.\end{aligned}$$

For the case of $j\neq i$,

$$\begin{aligned}\sum_{k=0}^{K}[\mathbf{G}_{k,i}+\mathbf{Q}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}&=\sum_{k=0}^{K}\sum_{r=0}^{k-1}\left[\lambda_i^r\lambda_j^{k-r}+\lambda_i^{r+1}\lambda_j^{k-r-1}\right]\mathbf{H}_k^\mathsf{T}\\&=\frac{\lambda_i+\lambda_j}{\lambda_i-\lambda_j}\left(\mathbf{H}(\lambda_i)-\mathbf{H}(\lambda_j)\right).\end{aligned}$$

Due to the integral Lipschitz hypothesis, both cases satisfy

$$\left\|\sum_{k=0}^{K}[\mathbf{G}_{k,i}+\mathbf{Q}_{k,i}]_{jj}\mathbf{H}_k^\mathsf{T}\right\|_2\leq 2C.$$

Then we obtain the upper bound of the summation of (16) and (17):

$$2C\sqrt{N}\epsilon^2\left\|\mathbf{X}\right\|_\mathsf{F}.\tag{26}$$

Saimilarily, the upper bound o Using the bound in (19) and (26), we obtain

$$\begin{aligned}&\left\|\sum_{k=0}^{K}\hat{\mathbf{S}}^k\mathbf{X}\mathbf{H}_k-\sum_{k=0}^{K}\mathbf{S}^k\mathbf{X}\mathbf{H}_k\right\|_\mathsf{F}\\&\quad\leq 2C\epsilon\left\|\mathbf{X}\right\|_\mathsf{F}+2C\sqrt{N}\epsilon^2\left\|\mathbf{X}\right\|_\mathsf{F}+\mathcal{O}\left(\epsilon^2\right)\\&\quad=2C\epsilon\left\|\mathbf{X}\right\|_\mathsf{F}+\mathcal{O}\left(\epsilon^2\right).\end{aligned}\tag{27}$$

■

VII. Proof Theorem 6

Proof: Let us consider the case where we have F_l features per layer, for $l=0,\ldots,L$. That is, the input to the GCNN is the collection of graph signals $\mathbf{X}_0=\mathbf{X}$ and the output $\mathbf{X}_0$ is a collection of F_L graph signals $\mathbf{X}_L$. In this context, we are interested in the difference between the output of the GCNNs when evaluated on different shift operators $\mathbf{S}$ and $\hat{\mathbf{S}}$. Firstly, we consider the the difference of layer ℓ

$$\begin{aligned}&\left\|\hat{\mathbf{X}}_\ell-\mathbf{X}_\ell\right\|_\mathsf{F}\\&\quad=\left\|\boldsymbol{\sigma}\left(\sum_{k=0}^{K}\hat{\mathbf{S}}^k\hat{\mathbf{X}}_{\ell-1}\mathbf{H}_{\ell k}\right)-\boldsymbol{\sigma}\left(\sum_{k=0}^{K}\mathbf{S}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}\right)\right\|_\mathsf{F}.\end{aligned}\tag{28}$$

Applying Lipschitz continuity of the nonlinearity, in which $|\sigma(b)-\sigma(a)|\leq|b-a|$, we get

$$\begin{aligned}&\left\|\hat{\mathbf{X}}_\ell-\mathbf{X}_\ell\right\|_\mathsf{F}\\&\quad\leq\left\|\sum_{k=0}^{K}\hat{\mathbf{S}}^k\hat{\mathbf{X}}_{\ell-1}\mathbf{H}_{\ell k}-\sum_{k=0}^{K}\mathbf{S}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}\right\|_\mathsf{F}.\end{aligned}\tag{29}$$

Adding and subtracting $\sum_{k=0}^{\infty}\hat{\mathbf{S}}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}$ from the terms in the sum, and using the triangular inequality, we get

$$\begin{aligned}&\left\|\sum_{k=0}^{K}\hat{\mathbf{S}}^k\hat{\mathbf{X}}_{\ell-1}\mathbf{H}_{lk}-\sum_{k=0}^{K}\mathbf{S}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}\right\|_\mathsf{F}\\&\leq\left\|\sum_{k=0}^{K}\hat{\mathbf{S}}^k\hat{\mathbf{X}}_{\ell-1}\mathbf{H}_{\ell k}-\sum_{k=0}^{K}\hat{\mathbf{S}}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}\right\|_\mathsf{F}\\&\quad+\left\|\sum_{k=0}^{K}\hat{\mathbf{S}}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}-\sum_{k=0}^{K}\mathbf{S}^k\mathbf{X}_{\ell-1}\mathbf{H}_{\ell k}\right\|_\mathsf{F}.\end{aligned}\tag{30}$$

By applying Lemma 4 on (**??**), we have that for $\forall \ell \geq 1$,

$$\left\|\hat{\mathbf{X}}_\ell - \mathbf{X}_\ell\right\|_{\mathsf{F}} \tag{31}$$

$$\leq B_\ell \left\|\hat{\mathbf{X}}_{\ell-1} - \mathbf{X}_{\ell-1}\right\|_{\mathsf{F}} + 2C_\ell \epsilon \left\|\mathbf{X}_{\ell-1}\right\|_{\mathsf{F}} + \mathcal{O}\left(\epsilon^2\right) \tag{32}$$

$$\leq B_\ell \left\|\hat{\mathbf{X}}_{\ell-1} - \mathbf{X}_{\ell-1}\right\|_{\mathsf{F}} + 2C_\ell \prod_{i=1}^{l-1} B_i \epsilon \left\|\mathbf{X}_0\right\|_{\mathsf{F}} + \mathcal{O}\left(\epsilon^2\right). \tag{33}$$

Solving the recursion bounds as in (31)-(33), we can bound the Frobenius norm of the output difference as

$$\left\|\hat{\mathbf{X}}_L - \mathbf{X}_L\right\|_{\mathsf{F}} \leq 2 \sum_{\ell=1}^{L} \frac{C_\ell}{B_\ell} \prod_{i=1}^{L} B_i \epsilon \left\|\mathbf{X}_0\right\|_{\mathsf{F}} + \mathcal{O}\left(\epsilon^2\right),$$

completing the proof. ■