Algebraic Neural Networks: Stability to Deformations

Alejandro Parada-Mayorga and Alejandro Ribeiro

***Abstract*—In this work we study the stability of *algebraic neural networks* (AlgNNs) with commutative algebras which unify CNNs and GNNs under the umbrella of algebraic signal processing. An AlgNN is a stacked layered structure where each layer is conformed by an algebra $\mathcal{A}$, a vector space $\mathcal{M}$ and a homomorphism $\rho : \mathcal{A} \to \text{End}(\mathcal{M})$, where $\text{End}(\mathcal{M})$ is the set of endomorphims of $\mathcal{M}$. Signals in each layer are modeled as elements of $\mathcal{M}$ and are processed by elements of $\text{End}(\mathcal{M})$ defined according to the structure of $\mathcal{A}$ via ρ. This framework provides a general scenario that covers several types of neural network architectures where formal convolution operators are being used. We obtain stability conditions regarding to perturbations which are defined as distortions of ρ, reaching general results whose particular cases are consistent with recent findings in the literature for CNNs and GNNs. We consider conditions on the domain of the homomorphisms in the algebra that lead to stable operators. Interestingly, we found that these conditions are related to the uniform boundedness of the Fréchet derivative of a function $p : \text{End}(\mathcal{M}) \to \text{End}(\mathcal{M})$ that maps the images of the generators of $\mathcal{A}$ on $\text{End}(\mathcal{M})$ into a power series representation that defines the filtering of elements in $\mathcal{M}$. Additionally, our results show that stability is universal to convolutional architectures whose algebraic signal model uses the same algebra.**

***Index Terms*—Algebraic Neural Networks, algebraic signal processing, representation theory of algebras, convolutional neural networks (CNNs), graph neural networks (GNNs), stability, Fréchet differentiability.**

I. Introduction

The numerical evidence that shows the benefits and goodness of using convolutional neural networks (CNNs) and graph neural networks (GNNs) for machine learning applications grows year after year, rising interest among researchers on finding solid and consistent explanations for the performance and the properties of these structures. In this context, the stability analysis of the operators representing the networks plays a central role, and some insights exist in the literature for CNNs [1]–[3] and GNNs [4]–[6], obtained independently but similar in form and nature. These results opened questions about whether these notions descend from a common notion of stability, and finding a connection or a general framework where these concepts can be unified have become of interest.

The notion of stability in CNNs is rooted in the notion of Lipschitz-continuity to the action of diffeomorphisms introduced by Mallat in [1] for the analysis of translation-invariant operators acting on $L^2(\mathbb{R}^n)$. This concept is adapted to invariant scattering convolutional networks [2], where the stability condition also implies global translation invariance. Mild variants of these notions have been considered in the literature for the analysis of convolutional networks. In [3] a version of the notion introduced in [1] is used considering the action of groups, formulating the Lipschitz stability in terms of the distance between a given group G and the invariance group G_0, while in [7] the stability condition is a variant of [1] considering signals whose domain is restricted to a compact subset of $\mathbb{R}^n$. For GNNs the problem of formulating stability conditions has been considered in [4], analyzing two scenarios, first the stability is measured with respect to changes in the amplitude of the signal, and second the perturbation is measured with respect to changes in the graphs. In [5] the notion of stability on graphs is considered in depth pointing out that the generalization of the conditions stated in [1], [2] is not straightforward for non smooth, non Euclidean domains, and as a way to quantify stability in GNNs the notion of *metric stability* is considered using a diffusion operator to measure the perturbations or changes in the graphs. In [6] a related notion of stability is used to provide concrete results about the stability on GNNs.

Although apparently different in nature, the stability results for CNNs and GNNs are uncannily similar not only in their final formulation but also in the analysis that lead to such results. Therefore, one may ask: is there any common framework where these proofs and analysis can be derived as particular cases?, can these stability results be extended to other convolutional architectures on different domains and with different notions of convolutions?. This paper addresses these questions, analyzing *algebraic neural networks (AlgNN)* to provide an affirmative answer. In particular, we analyze the stability properties of AlgNNs deriving general results whose particular instances lead to those results known for GNNs and CNNs. An AlgNN is a layered structure where the processing of the data in each layer is done in the context of an *algebraic signal model* in the sense of [8]. In particular, each layer is defined by an algebra $\mathcal{A}$, a vector space $\mathcal{M}$ where the signals, as elements of $\mathcal{M}$, are processed by elements of the set of endomorphisms $\text{End}(\mathcal{M})$, which are selected according to an homomorphism $\rho : \mathcal{A} \to \text{End}(\mathcal{M})$. Using the map that relates an operator and its perturbed version, we introduce a notion of stability that uses the Fréchet derivative of this map and we provide stability bounds which depend on the algebraic structure of the signal models and whose particular instantiations lead to results like the one provided in [6]. For this analysis we consider a perturbation model that contains the already studied perturbation models analyzed in [4]–[6]. We remark as the most important contributions of this paper:

- The introduction of AlgNNs to unify stability analysis of neural networks where formal convolutions are used. In particular, we introduce the notion of perturbations and stability to algebraic signal models (definition 5), where the size of the perturbation is measured taking into account the Fréchet derivative of an operator that relates a filter with its perturbed version. This notion is compatible with the notion of stability considered in [1]–[6].
- Stability results for algebraic filters considering commutative algebras (theorems 1 to 4 and corollaries 1 and 2). As a consequence of these results we obtain also stability results for AlgNNs (theorems 5 and 6), whose particular instantiations lead to known results for CNNs and GNNs.

The results presented in theorems 1 to 4 and corollaries 1 and 2 have several important implications among which are the following. First, they show that the stability properties in

Department of Electrical and Systems Engineering, University of Pennsylvania. Email: alejopm@seas.upenn.edu, aribeiro@seas.upenn.edu.

convolutional architectures can be expressed in terms of the algebraic laws (structure of $\mathcal{A}$) that govern the signal model in each layer. This fact highlights that the property of stability is universal always that the algebra remains the same, and it also explains the remarkable similarities between the results obtained for GNNs and CNNs. Second, the changes in the filters acting on a given signal are determined by the Fréchet derivative of the filters as functions of the images of the generators in the algebra, and this is true no matter which algebra is used and what particular representation is being considered. At the same time, this derivative as an operator acting on the perturbation determines whether the filters are stable or not. Third, although the perturbation is considered on the images of the algebra on the space of endomorphisms, the stability to a given perturbation is determined by the existence of certain subsets of the algebra itself, i.e. the stability relies on restrictions that are imposed on the filters and as a consequence the AlgNN can resolve the frequency selectivity without loosing stability. We remark that these ideas contribute to a deeper understanding of convolutional architectures and explain behaviors, results and analysis that exhibit similarities but that were conceived by different means and tools due to an apparent disconnection between them, for instance like with CNNs and GNNs.

This paper is organized as follows. In Section II we start discussing the basics about algebraic signal processing and algebraic filtering introducing the basics about representation theory of algebras and its connection to the general algebraic signal model, we provide a set of examples to make basic concepts and ideas clear to the reader. In Section III we introduce algebraic neural networks discussing how this notion generalizes CNNs, Group-CNNs, Graphon-CNNs, and GNNs which are considered as particular instantiations. The notions of perturbations and stability are discussed in Section IV where the formal definition of stability is introduced with the perturbation model considered for our discussion. Additionally, several stability theorems are presented in Section V considering filters in the algebraic signal model and operators representing AlgNNs as well. In Section VI we discuss the basics about the spectral representation of algebraic filters, relevant to Section VII where the proofs of the theorems in Section IV are presented. Section VIII offers a discussion about the results pointing out the role of subsets of the algebra and the stability to perturbations given a perturbation model. Finally, some conclusions are provided in Section IX.

II. Algebraic Filters

Algebraic signal processing (ASP) provides a framework for understanding and generalizing traditional signal processing exploiting the representation theory of algebras [8]–[11]. In ASP, a signal model is defined as the triple

$$(\mathcal{A}, \mathcal{M}, \rho), \tag{1}$$

in which $\mathcal{A}$ is an associative algebra with unity, $\mathcal{M}$ is a vector space with inner product, and $\rho : \mathcal{A} \to \text{End}(\mathcal{M})$ is a homomorphism between the algebra $\mathcal{A}$ and the set of endomorphisms in the vector space $\mathcal{M}$. The elements in (1) are tied together by the notion of a representation which we formally define next. In this paper we further restrict attention to commutative algebras $\mathcal{A}$.

Definition 1 (Representation)**.** *A representation of the associative algebra $\mathcal{A}$ is a vector space $\mathcal{M}$ equipped with a homomorphism $\rho : \mathcal{A} \to End(\mathcal{M})$, i.e., a linear map preserving multiplication and unit.*

When denoting a representation of a given algebra we will use a pair consisting of the vector space and the homomorphism, for instance with the notation in definition 1 the representation of the algebra $\mathcal{A}$ is denoted by $(\mathcal{M}, \rho)$. The class of representations of an algebra $\mathcal{A}$ is denoted as $\mathsf{Rep}\{\mathcal{A}\}$.

In an ASP model, signals are elements of the vector space $\mathcal{M}$, and filters are elements of the algebra $\mathcal{A}$. Thus, the vector space $\mathcal{M}$ deteremines the objects of interest and the algebra $\mathcal{A}$ the rules of the operations that define a filter. The homomorphism ρ translates the abstract operators $a \in \mathcal{A}$ into concrete operators $\rho(a)$ that act on signals $\mathbf{x}$ to produce filter outputs

$$\mathbf{y} = \rho(a)\mathbf{x}. \tag{2}$$

The algebraic filters in (2) generalize the convolutional processing of time signals – see Example 1. Our goal in this paper is to use them to generalize convolutional neural networks (Section III) and to study their fundamental stability properties (Section IV). Generators, which we formally define next, are important for the latter goal.

Definition 2 (Generators)**.** *For an associative algebra with unity $\mathcal{A}$ we say the set $\mathcal{G} \subseteq \mathcal{A}$ generates $\mathcal{A}$ if all $a \in \mathcal{A}$ can be represented as polynomial functions of the elements of $\mathcal{G}$. We say elements $g \in \mathcal{G}$ are generators of $\mathcal{A}$ and we denote as $a = p_{\mathcal{A}}(\mathcal{G})$ the polynomial that generates a.*

Definition 2 states that elements $a \in \mathcal{A}$ can be built from the generating set as polynomials using the operations of the algebra. Given the role of representations in connecting the algebra $\mathcal{A}$ with the signals $\mathbf{x}$ as per Definition 1 the representation $\rho(g)$ of a generator $g \in \mathcal{G}$ will be of interest. In the context of ASP, these representations are called shift operators as we formally define next.

Definition 3 (Shift Operators)**.** *Let $(\mathcal{M}, \rho)$ be a representation of the algebra $\mathcal{A}$. Then, if $\mathcal{G} \subset \mathcal{A}$ is a generator set of $\mathcal{A}$, the operators $\mathbf{S} = \rho(g)$ with $g \in \mathcal{G}$ are called shift operators. The set of all shift operators is denoted by $\mathcal{S}$*

Given that elements a of the algebra are generated from elements g of the generating set, it follows that filters $\rho(a)$ are generated from the set of shift operators $\mathbf{S} = \rho(g)$. In fact, if we have that $a = p_{\mathcal{A}}(\mathcal{G})$ is the polynomial that generates a, the filter $\rho(a)$ is generated by

$$\rho(a) = p_{\mathcal{M}}\big(\rho(\mathcal{G})\big) = p_{\mathcal{M}}\big(\mathcal{S}\big) = p\big(\mathcal{S}\big), \tag{3}$$

where the subindex $\mathcal{M}$ signifies that the operations in (3) are those of the vector space $\mathcal{M}$ – in contrast to the polynomial $a = p_{\mathcal{A}}(\mathcal{G})$ whose operations are those of the algebra $\mathcal{A}$. In the last equality we drop the subindex $\mathcal{M}$ to simplify notation as it is generally understood from context.

To clarify ideas it is instructive to consider some examples.

Example 1 (Discrete Time Signal Processing)**.** Let $\mathcal{M} = \ell^2$ be the space of square summable sequences $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}}$ and $\mathcal{A}$ the algebra of polynomials with elements $a = \sum_{k=0}^{K-1} h_k t^k$. We consider the shift operator S such that $S\mathbf{x}$ is the sequence with entries

$(S\mathbf{x})_n = x_{n-1}$. With the homomorphism $\rho(a) = \sum_{k=0}^{K-1} h_k S^k$ the filters in (2) are of the form

$$\mathbf{y} = \rho\left(\sum_{k=0}^{K-1} h_k t^k\right)\mathbf{x} = \left(\sum_{k=0}^{K-1} h_k S^k\right)\mathbf{x} = \sum_{k=0}^{K-1} h_k S^k \mathbf{x}. \quad (4)$$

This is a representation of time convolutional filters [8]. Observe that the algebra of polynomials is generated by $g = t$ and that since $\rho(t) = S$ the shift operator S generates the set of convolutional filters as it follows from (4).

Example 2 (Graph Signal Processing)**.** We retain the algebra of polynomials as in Example 1 but we change the space of signals to the set of complex vectors with N entries, $\mathcal{M} = \mathbb{C}^N$. We interpret components x_n of $\mathbf{x} \in \mathcal{M} = \mathbb{C}^N$ as being associated with nodes of a graph with matrix representation $\mathbf{S} \in \mathbb{C}^{N\times N}$. With the homomorphism $\rho(a) = \sum_{k=0}^{K-1} h_k \mathbf{S}^k$ the filters in (2) are of the form

$$\mathbf{y} = \rho\left(\sum_{k=0}^{K-1} h_k t^k\right)\mathbf{x} = \left(\sum_{k=0}^{K-1} h_k \mathbf{S}^k\right)\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}. \quad (5)$$

This is a representation of the graph convolutional filters used in graph signal processing (GSP) [12], [13]. Observe that (5) and (4) are similar but represent different operations. In (5) $\mathbf{x}$ is a vector and $\mathbf{S}^k$ a matrix power. In (4) $\mathbf{x}$ is a sequence and S^k is the composition of the shift operator S. Their similarity arises from their common use of the algebra of polynomials. Their differences are because they use different vector spaces $\mathcal{M}$ and different homomorphism functions ρ.

Example 3 (Discrete Signal Processing)**.** Make $\mathcal{M} = \mathbb{C}^N$ and $\mathcal{A}$ the algebra of polynomials in the variable t and modulo $t^N - 1$. Elements of $\mathcal{A}$ are of the form $a = \sum_{k=0}^{K-1} h_k t^k$ but we must have $K \leq N$ and polynomial products use the rule $t^k = t^{k \mod N}$. Define the cyclic shift operator C as one whose application to $\mathbf{x} \in \mathcal{M} = \mathbb{C}^N$ yields a vector with components $(C\mathbf{x})_n = x_{(n-1) \mod N}$. With the homomorphism $\rho(a) = \sum_{k=0}^{K-1} h_k C^k$ the filters in (2) are of the form

$$\mathbf{y} = \rho\left(\sum_{k=0}^{K-1} h_k t^k\right)\mathbf{x} = \left(\sum_{k=0}^{K-1} h_k C^k\right)\mathbf{x} = \sum_{k=0}^{K-1} h_k C^k \mathbf{x}. \quad (6)$$

This is the representation of (cyclic) convolutional filters in discrete signal processing [8]. We emphasize that the homomorphism $\rho(a) = \sum_{k=0}^{K-1} h_k C^k$ is indeed a homomorphism because the cyclic shift operator C satisfies $C^k = C^{k \mod N}$.

As is clear from Examples 1-3, the effect of the operator $\rho(a)$ on a given signal $\mathbf{x}$ is determined by two factors: The filter $a \in \mathcal{A}$ and the homomorphism ρ. The filter $a \in \mathcal{A}$ indicates the *laws and rules* to be used to manipulate the signal $\mathbf{x}$ and ρ provides a *physical realization* of the filter a on the space $\mathcal{M}$ to which $\mathbf{x}$ belongs. For instance, in these three examples the filter $a = 1 + 2t$ indicates that the signal is to be added to a transformed version of the signal scaled by coefficient 2. The homomorphism ρ in Example 1 dictates that the physical implementation of this transformation is a time shift. The homomorphism ρ in Example 2 defines a transformation as a multiplication by $\mathbf{S}$ and in Example 3 the homomorphism entails a cyclic shift. We remark that in order to specify the physical effect of a filter it is always sufficient to specify the physical effect of the generators. In all three examples, the generator of the algebra is $g = t$. The respective effects of an arbitrary filter a are determined once we specify that $\rho(t) = S$ in Example 1, $\rho(t) = \mathbf{S}$ in Example 2, or $\rho(t) = C$ in Example 3.

The flexibility in the choice of algebra and homomorphism allows for a very rich variety of signal processing frameworks. We highlight this richness with three more examples.

Example 4 (Signal Processing on Groups)**.** Let $\mathcal{M} = \{\mathbf{x} : G \to \mathbb{C}\} = \{\sum_{g\in G} \mathbf{x}(g)g\}$ be the set of functions defined on the group G with values in $\mathbb{C}$ and $\mathcal{A} = \mathcal{M}$ the *group algebra* with elements $\sum_{g\in G} \mathbf{a}(g)g$. With a homomorphism given by $\rho(\mathbf{a}) = \mathbf{a}$ the filtering in (2) takes the form

$$\rho\left(\sum_{g\in G} \mathbf{a}(g)g\right)\mathbf{x} = \sum_{g\in G} \mathbf{a}(g)g\mathbf{x} = \sum_{g\in G}\sum_{h\in G} \mathbf{a}(g)\mathbf{x}(h)gh, \quad (7)$$

and making $u = gh$ we have

$$\sum_{g,h\in G} \mathbf{a}(g)\mathbf{x}(h)gh = \sum_{u,h\in G} \mathbf{a}(uh^{-1})\mathbf{x}(h)u = \mathbf{a} * \mathbf{x}. \quad (8)$$

This is the standard representation of convolution of signals on groups [14]–[16]. We point out that (7) and (8) hold for any group but that not all group algebras are commutative. Results in Section V apply only when the group group algebra is commutative.

Example 5 (Graphon Signal Processing)**.** Let $\mathcal{M} = L^2([0,1])$ be the space of signals and $\mathcal{A}$ the algebra of polynomials with elements $a = \sum_{k=0}^{K-1} h_k t^k$ with a homomorphism given by $\rho(a) = \sum_{k=0}^{K-1} h_k S^k$ where

$$(S\mathbf{x})(u) = \int_0^1 W(u,v)\mathbf{x}(v)dv, \quad (9)$$

where $W(u,v) : [0,1]^2 \to [0,1]$ is a graphon, i.e. bounded symmetric measurable function [17], [18]. Then, the filtering in (2) is of the form (4). This is a representation of convolutional filters in the context of graphon signal processing [19].

The choice of $\mathcal{A}$ and ρ provides means to leverage our knowledge of the signal's domain in its processing. The convolutional filters in (4) leverage the shift invariance of time signals and the filters in (6) the cyclic invariance of periodic signals. The group convolutional filters in (8) generalize shift invariance to invariance with respect to an arbitrary group action. The graph convolutional filters in (5) engender signal processing that is independent of node labeling [20] and the graphon filters in Example 5 a generalization of this notion to dense domains [19]. Leveraging this structure is instrumental in achieving scalable information processing. In the following section we explain how neural network architectures combine algebraic filters as defined in (2) with pointwise nonlinearities to attain signal processing that inherits the invariance properties of the respective algebraic filters.

III. Algebraic Neural Networks

With the concept of algebraic filtering at hand we define an algebraic neural network (AlgNN) as a stacked layered structure (see Fig. 1) in which each layer is composed by the triple $(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)$, which is an algebraic signal model associated to each layer. Notice that $(\mathcal{M}_\ell, \rho_\ell)$ is a representation of $\mathcal{A}_\ell$. The mapping between layers is performed by the maps $\sigma_\ell : \mathcal{M}_\ell \to \mathcal{M}_{\ell+1}$ that perform those operations of point-wise nonlinearity

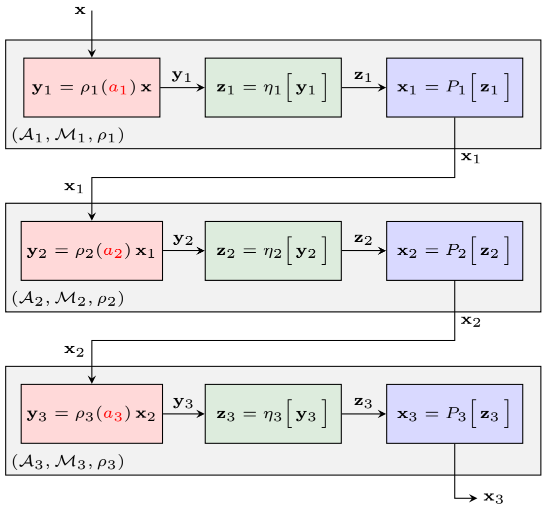

Figure 1. Algebraic Neural Network $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^{3}$ with three layers indicating how the input signal $\mathbf{x}$ is processed by Ξ and mapped into $\mathbf{x}_3$.

and pooling. Then, the ouput from the layer ℓ in the AlgNN is given by

$$\mathbf{x}_\ell = \sigma_\ell\left(\rho_\ell(a_\ell)\mathbf{x}_{\ell-1}\right) \tag{10}$$

where $a_\ell \in \mathcal{A}_\ell$, which can be represented equivalently as

$$\mathbf{x}_\ell = \Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_{\ell-1}, \mathcal{S}_{\ell-1}), \tag{11}$$

where $\mathcal{P}_\ell \subset \mathcal{A}_\ell$ highlights the properties of the filters and $\mathcal{S}_\ell$ is the set of shifts associated to $(\mathcal{M}_\ell, \rho_\ell)$. Additionally, the term $\Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L\right)$ represents the total map associated to an AlgNN acting on a signal $\mathbf{x}$.

1) Convolutional Features: The processing in each layer can be performed by means of several families of filters, which will lead to several *features*. In particular the feature f obtained in the layer ℓ is given by

$$\mathbf{x}_\ell^f = \sigma_\ell\left(\sum_{g=1}^{F_\ell} \rho_\ell\left(a_\ell^{gf}\right)\mathbf{x}_{\ell-1}^g\right), \tag{12}$$

where a_ℓ^{fg} is the filter in $\mathcal{A}_\ell$ used to process the g-th feature $\mathbf{x}_{\ell-1}^g$ obtained from layer $\ell-1$ and F_ℓ is the number of features.

2) Pooling: As stated in [21] the pooling operation in CNNs helps to keep representations approximately invariant to small translations of an input signal, and also helps to improve the computational efficiency. In this work this operation is attributed to the operator σ_ℓ. In particular, we consider $\sigma_\ell = P_\ell \circ \eta_\ell$ where P_ℓ is a pooling operator and η_ℓ is a pointwise nonlinearity. The only property assumed from σ_ℓ is to be Lipschitz and to have zero as a fixed point, i.e. $\sigma_\ell(0) = 0$. It is important to point out that P_ℓ projects elements from a given vector space into another.

Example 6 (Traditional CNNs)**.** Traditional CNNs rely on the use of typical signal processing models and can be considered a particular case of an AlgNN where the algebraic signal model is the same as in example 3. Consequently, the fth feature in layer ℓ is given by

$$\mathbf{x}_\ell^f = \sigma_\ell\left(\sum_{g=1}^{F_\ell}\sum_{i=1}^{K} h_i^{gf}\mathbf{S}_\ell^i\mathbf{x}_{\ell-1}^g\right), \tag{13}$$

where $\rho_\ell(t) = \mathbf{S}_\ell$. In this case P_ℓ is a sampling operator while typically $\eta_\ell(u) = \max\{0, u\}$.

Example 7 (Graphon Neural Networks)**.** In graphon neural networks the algebraic signal model in each layer corresponds to the one discussed in example 5. Consequently, the fth feature in layer ℓ has the form given in eqn. (13) considering a shift S given according to eqn. (9). It is worth noticing that in this case the signals in the inner layers of the network are meant to be represented using a finite number of eigenvectors of the shift operator. Then, if we denote by $\boldsymbol{\varphi}_i^{(\ell)}$ the eigenvectors of the graphon $W_\ell(u,v)$ associated to the ℓth layer P_ℓ is given by

$$P_\ell : \text{span}\left\{\boldsymbol{\varphi}_i^{(\ell)}\right\}_{i\in I} \to \text{span}\left\{\boldsymbol{\varphi}_j^{(\ell+1)}\right\}_{j\in J}, \tag{14}$$

where I and J are finite subsets of $\mathbb{N}$. Then, the signals in each $\mathcal{M}_\ell$ are always functions of a continuous independent variable but with *lower degrees of freedom* as ℓ increases.

Example 8 (Group Neural Networks)**.** In group neural networks the algebraic model is the same as specified in example 4. Therefore, the fth feature in layer ℓ is given by

$$\mathbf{x}_\ell^f = \sigma_\ell\left(\sum_{n=1}^{F_\ell}\sum_{u,h\in G_\ell}\mathbf{a}_{(\ell)}^{nf}(uh^{-1})\mathbf{x}_{\ell-1}(h)u\right). \tag{15}$$

where G_ℓ is the group associated to the ℓth layer and $\mathbf{a}_{(\ell)}^{nf}$ are the coefficients of the filter associated to the feature f in layer ℓ. In this case $P_\ell : L^2(G_\ell) \to L^2(G_{\ell+1})$, where $L^2(G)$ is the set of signals of finite energy defined on the group G. If the groups G_ℓ are finite P_ℓ can be conceived as a typical projection mapping between $\mathbb{R}^{|G_\ell|} \to \mathbb{R}^{|G_{\ell+1}|}$.

IV. Perturbations

In an ASP triple $(\mathcal{A}, \mathcal{M}, \rho)$, signals $\mathbf{x} \in \mathcal{M}$ are observations of interest and the algebra $\mathcal{A}$ defines the operations that are to be performed on signals. The homomorphism ρ ties these two objects and, as such, is one we can consider as subject to model mismatch. In this paper we consider perturbations adhering to the following model.

Definition 4. *(ASP Model Perturbation) Let $(\mathcal{A}, \mathcal{M}, \rho)$ be an ASP model with algebra elements generated by $g \in \mathcal{G}$ (Definition 2) and recall the definition of the shift operators $\mathbf{S} = \rho(g)$ (Definition 3). We say that $(\mathcal{A}, \mathcal{M}, \tilde{\rho})$ is a perturbed ASP model if for all $a = p_{\mathcal{A}}(\mathcal{G})$ we have that*

$$\tilde{\rho}(a) = p_{\mathcal{M}}\big(\tilde{\rho}(g)\big) = p_{\mathcal{M}}\big(\tilde{\mathcal{S}}\big) = p\big(\tilde{\mathcal{S}}\big), \tag{16}$$

where $\tilde{\mathcal{S}}$ is a set of perturbed shift operators of the form

$$\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{T}(\mathbf{S}), \tag{17}$$

for all shift operators $\mathbf{S} \in \mathcal{S}$

As per Definition 4, an ASP perturbation model, is a perturbation of the homomorphism ρ defined by a perturbation of the shift operators $\mathbf{S}$. Each shift operator $\mathbf{S}$ is perturbed to the shift operator $\tilde{\mathbf{S}}$ according to (17) and this perturbation propagates to the filter $\rho(a)$ according to (16). An important technical remark is that the resulting mapping $\tilde{\rho}$ is not required to be a homomorphism – although it can be, indeed, often is.

We point out that Definition 4 limits the perturbation of the homomorphism ρ to perturbations of the shift operators. This is a somewhat arbitrary choice but one that is motivated by practice as we illustrate for graph signals (Example 9), discrete time signals (Example 10), signals on groups (Example 12) and graphon signals (Example 11). The model is also arbitrary in that it perturbs the homomorphism ρ but not the Algebra $\mathcal{A}$ or the signal $\mathbf{x}$. This is also justifiable. Notice, first that the algebra $\mathcal{A}$ is a choice of operations and therefore not naturally subject to perturbation. Perturbing signals $\mathbf{x}$ is of interest but a transformation of $\mathbf{x}$ can be reinterpreted as a transformation of $\mathbf{S}$. Indeed, if we are given signals $\mathbf{x}$ and $T\mathbf{x}$ we can always define a shift operator $\tilde{\mathbf{S}} = \mathbf{S}T$ and write

$$\mathbf{S}(T\mathbf{x}) = (\mathbf{S}T)\mathbf{x} = \tilde{\mathbf{S}}\mathbf{x}. \tag{18}$$

Thus, we understand the effect of the perturbation $T\mathbf{x}$ by studying the processing of $\mathbf{x}$ with the perturbed shift operator $\tilde{\mathbf{S}} = \mathbf{S}T$. We discuss this further for discrete time signals in Example 10 and for signals on groups in Example 12 in the following. We also discuss the model in (4) particularized to graphs (Example 9) and graphons (Example 11).

A. Perturbation Models

In our subsequent analysis we consider perturbation models of the form

$$\mathbf{T}(\mathbf{S}) = \mathbf{T}_0 + \mathbf{T}_1\mathbf{S}, \tag{19}$$

which is a generic model of small perturbations of a shift operator that involve an absolute perturbation $\mathbf{T}_0$ and a relative perturbation $\mathbf{T}_1\mathbf{S}$; see [6]. The $\mathbf{T}_r$ are compact normal operators with operator norm $\|\mathbf{T}_r\| < 1$. Requiring $\|\mathbf{T}_r\| < 1$ is a minor restriction as we are interested in small perturbations with $\|\mathbf{T}_r\| \ll 1$.

For the model in (19) it is important to describe the commutativity of the shift operator $\mathbf{S}$ and the perturbation model matrices $\mathbf{T}_r$. To that end, we write

$$\mathbf{S}\mathbf{T}_r = \mathbf{T}_{cr}\mathbf{S} + \mathbf{S}\mathbf{P}_r, \tag{20}$$

where $\mathbf{T}_{cr} = \sum_i \mu_i \mathbf{u}_i \langle \mathbf{u}_i, \cdot \rangle$, μ_i is the ith eigenvalue of $\mathbf{T}_r$, $\mathbf{u}_i$ is the ith eigenvector of $\mathbf{S}$, and $\langle , \rangle$ represents the inner product operation. As a consequence, we have that $\mathbf{S}\mathbf{T}_{rc} = \mathbf{T}_{rc}\mathbf{S}$ and $\|\mathbf{T}_{cr}\| = \|\mathbf{T}_r\|$. We define the commutation factor

$$\delta = \max_r \frac{\|\mathbf{P}_r\|}{\|\mathbf{T}_r\|}, \tag{21}$$

which is a measure of how far the operators $\mathbf{S}$ and $\mathbf{T}_r$ are from commuting with each other. Notice that $\delta = 0$ implies $\mathbf{P}_r = \mathbf{0}$ and $\mathbf{T}_r = \mathbf{T}_{cr}$. The commutation factor δ in (21) can be bounded in terms of the spectra of $\mathbf{T}_r$ and $\mathbf{S}$ as we show in Proposition 2. The specifics of this bound are not central to the results of Section V.

Remark 1. Notice that δ is a property of the perturbation and not from the filters (see remark 3). As we will show in the next section a large value of δ does not modify the property of being stable, although it increases the values of the constants associated to stability. Studying strategies to reduce the effects of δ in the stability constants is an interesting future research direction.

Notice that when representations of an algebra $\mathcal{A}$ with multiple generators $\{g_i\}_{i=1}^m$ are considered, we have that for $a \in \mathcal{A}$ the operator $p(\rho(a)) \in \text{End}(\mathcal{M})$ is a function of $\rho(g_i) = \mathbf{S}_i \in \text{End}(\mathcal{M})$ and therefore can be seen as the function $p : \text{End}(\mathcal{M})^m \to \text{End}(\mathcal{M})$, where $\text{End}(\mathcal{M})^m$ is the m-times cartesian product of $\text{End}(\mathcal{M})$. In this scenario we use the notation $p(\mathbf{S}) = p(\mathbf{S}_1, \ldots, \mathbf{S}_m)$ and when considering the perturbation model in eqn. (19) acting on $\mathbf{S} = (\mathbf{S}_1, \ldots, \mathbf{S}_m)$ we use the following notation $\mathbf{T}(\mathbf{S}) = (\mathbf{T}(\mathbf{S}_1), \ldots, \mathbf{T}(\mathbf{S}_m))$ where $\mathbf{T}(\mathbf{S}_i) = \mathbf{T}_{0,i} + \mathbf{T}_{1,i}\mathbf{S}_i$.

V. Stability Theorems

The filters in Section II and the algebraic neural networks in Section III are operators acting on the space $\mathcal{M}$. These operators are of the form $p(\mathbf{S})$, and their outputs depend on a filter set $\mathcal{H} \subset \mathcal{A}$ which is denoted as $p_{\mathcal{A}} \in \mathcal{H} \subset \mathcal{A}$, and the set of shift operators $\mathcal{S}$, where $\mathbf{S} \in \mathcal{S}$. When we perturb the processing model according to Definition 4, these operators are perturbed as well. The goal of this paper is to analyze these perturbations. In particular, our goal is to identify conditions for filters and algebraic neural networks to be stable in the sense of the following definition.

Definition 5 (Operator Stability)**.** *Given operators $p(\mathbf{S})$ and $p(\tilde{\mathbf{S}})$ defined on the processing models $(\mathcal{A}, \mathcal{M}, \rho)$ and $(\mathcal{A}, \mathcal{M}, \tilde{\rho})$ (cf. Definition 4) we say the operator $p(\mathbf{S})$ is Lipschitz stable if there exist constants $C_0, C_1 > 0$ such that*

$$\left\| p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x} \right\| \leq \left[C_0 \sup_{\mathbf{S} \in \mathcal{S}} \|\mathbf{T}(\mathbf{S})\| + C_1 \sup_{\mathbf{S} \in \mathcal{S}} \left\| D_{\mathbf{T}}(\mathbf{S}) \right\| + \mathcal{O}\left(\|\mathbf{T}(\mathbf{S})\|^2 \right) \right] \|\mathbf{x}\|, \tag{22}$$

for all $\mathbf{x} \in \mathcal{M}$. In (22) $D_{\mathbf{T}}(\mathbf{S})$ is the Fréchet derivative of the perturbation operator $\mathbf{T}$.

When the perturbation value $\mathbf{T}(\mathbf{S})$ and its derivative $D_{\mathbf{T}}(\mathbf{S})$ are small, the inequality in (22) states that the operators $p(\mathbf{S})$ and $p(\tilde{\mathbf{S}})$ are close uniformly across all inputs $\mathbf{x}$. Our stability theorems are presented in the next section, but at this point it is important to remark that algebraic filters are not always stable in the sense of (22). We know that this is true because unstable counterexamples are known in the case of graph signal processing [20] and the processing of time signals [1].

Example 9 (Graph Signal Processing)**.** In graph signal processing the shift operator $\mathbf{S}$ is a matrix representation of a graph. The perturbation model in (4) simply states that $\tilde{\mathbf{S}}$ is a matrix representation of a different graph. Definition 5 defines a stable operator $p(\mathbf{S})$ as one that doesn't change much when run on graphs that are close and related by perturbations that are sufficiently smooth in the space of matrix representations of graphs. The notion of absolute graph perturbation considered for the stability analysis of GNNs in [6] is a particular case of eqn. (19) with $\mathbf{T}_1 = \mathbf{0}$. Additionally, the notion of relative graph perturbations can be obtained from eqn. (19) considering $\mathbf{T}_0 = \mathbf{0}$.

Example 10 (Discrete Time Signal Processing)**.** In discrete time signal processing the shift operator is the discrete time shift S. In this case the perturbation model is more meaningful if interpreted as a perturbation of the signal $\mathbf{x}$. The processing induced by (4) is invariant to shifts and therefore adequate to processing signals that are shift invariant. In general, signals are close to shift invariant but not exactly so. That is, a given signal $\mathbf{x}$ is invariant with respect to a shift operator $\tilde{S}$ that is close to the time shift S. If the stability property in (22) holds we can guarantee that processing the signal $\mathbf{x}$ with the operator $\tilde{S}$ is not far from processing the signal with the operator S. The latter represents the operations we perform – since we choose to use S in the processing of time signals. The former represents the processing we should undertake to respect the actual invariance properties of the signal $\mathbf{x}$ – which are characterized by $\tilde{S}$.

Example 11 (Graphon Signal Processing)**.** Similar to Example 9 the graphon $W(u,v)$ is a limit object that represents a family of random graphs. The perturbed graphon $\widetilde{W}(u,v)$ represents a different family of random graphs. The perturbation of the graphon generates a corresponding perturbation of the shift operator defined in (9). If the condition stated in (22) is satisfied, then filtering graphon signals using the perturbed shift operator associated to $\widetilde{W}(u,v)$ will lead to similar results to the ones obtained with the unperturbed operator and the differences are proportional to the size of the perturbation acting on $W(u,v)$.

Example 12 (Group Signal Processing)**.** Similar to Example 10, the filters in (7) are invariant to the action of the group. Actual signals $\mathbf{x}$ are invariant to actions of operators that are close to actions of the group – e.g., a signal is close to invariant to rotations and symmetries. If (22) is true, processing the signal with operators g – as we choose to do – is not far from processing the signal with operators $\tilde{g}$ – as we should do to leverage the actual invariance of the signal $\mathbf{x}$. We remark that when we perturb g to $\tilde{g}$ the resulting shift operators will not, in general, be representations of a homomorphism.

A. Stability of Algebraic Filters

Taking into account that the notion of stability is meant to be satisfied by subsets of filters of the algebra and not necessarily the whole algebra, it is important to have a characterization of these subsets in simple terms. In the case of commutative algebras with m generators this characterization can be stated using m-multivariate functions. To do so, we use an algebra $\mathcal{A}_{\mathbb{C}}$ isomorphic to $\mathcal{A}$, where the symbols are variables taking values in $\mathbb{C}$ and then define conditions in this new algebra. For instance, if $\mathcal{A}$ is the algebra generated by g with coefficients in $\mathbb{R}$, we have the isomorphism $\iota : \mathcal{A} \to \mathcal{A}_{\mathbb{C}}$ given by $\iota(g) = \lambda$ where $\lambda \in \mathbb{C}$. Then, $\iota\left(p_{\mathcal{A}}(g)\right) = p(\lambda)$, and stating conditions on $p(\lambda)$ we can define subsets of $\mathcal{A}$ via ι^{-1}. In the following paragraphs we introduce some definitions used to characterize subsets of algebras with one generator. Results associated to multiple generators will be discussed in the next subsection.

Definition 6. *Let $p : \mathbb{C} \to \mathbb{C}$ be single variable function. Then, it is said that p is Lipschitz if there exists $L_0 > 0$ such that*

$$|p(\lambda) - p(\mu)| \leq L_0 |\lambda - \mu| \tag{23}$$

for all $\lambda, \mu \in \mathbb{C}$. Additionally, it is said that $p(\lambda)$ is Lipschitz integral if there exists $L_1 > 0$ such that

$$\left| \lambda \frac{dp(\lambda)}{d\lambda} \right| \leq L_1 \text{ for all } \lambda. \tag{24}$$

In what follows, when considering subsets of a commutative algebra $\mathcal{A}$, we denote by $\mathcal{A}_{L_0}$ the subset of elements in $\mathcal{A}$ that are Lipschitz with constant L_0 and by $\mathcal{A}_{L_1}$ the subset of element of $\mathcal{A}$ that are Lipschitz integral with constant L_1. Additionally, for the sake of simplicity we will not make reference to ι.

We start with a result for operators in algebraic models with a single generator, highlighting the role of the Fréchet derivative of the map that relates the operator and its perturbed version.

Theorem 1. *Let $\mathcal{A}$ be an algebra generated by g and let $(\mathcal{M}, \rho)$ be a representation of $\mathcal{A}$ with $\rho(g) = \mathbf{S} \in End(\mathcal{M})$. Let $\tilde{\rho}(g) = \tilde{\mathbf{S}} \in End(\mathcal{M})$ where the pair $(\mathcal{M}, \tilde{\rho})$ is a perturbed version of $(\mathcal{M}, \rho)$ and $\tilde{\mathbf{S}}$ is related to $\mathbf{S}$ by the perturbation model in eqn. (17). Then, for any $p_{\mathcal{A}} \in \mathcal{A}$ we have*

$$\left\| p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x} \right\| \leq \|\mathbf{x}\| \left(\|D_p(\mathbf{S})\left\{\mathbf{T}(\mathbf{S})\right\}\| + \mathcal{O}\left(\|\mathbf{T}(\mathbf{S})\|^2\right) \right) \tag{25}$$

where $D_p(\mathbf{S})$ is the Fréchet derivative of p on $\mathbf{S}$.

Proof. See Section VII-A □

Theorem 1 highlights an important point, the difference between two operators obtained from the same elements in the algebra is bounded by the Fréchet derivative of $p(\mathbf{S})$ which depends of the properties of the elements in $\mathcal{A}$. In particular, we can see that an upper bound in the term $\|D_p(\mathbf{S})\mathbf{T}(\mathbf{S})\|$ depends on how the the operator $D_p(\mathbf{S})$ acts on the perturbation $\mathbf{T}(\mathbf{S})$. Then, $D_p(\mathbf{S})$ will determine whether $p(\mathbf{S})$ is stable under the effect of $\mathbf{T}(\mathbf{S})$, or in other words the properties of $p_{\mathcal{A}}$ act on the perturbation via the operator $D_p(\mathbf{S})$. Additionally, notice that eqn. (25) is satisfied for any $\mathbf{T}(\mathbf{S})$ if $D_p(\mathbf{S})$ exists.

In the following theorems we show how these terms are related to $\mathbf{T}(\mathbf{S})$ and its Fréchet derivative $D_{\mathbf{T}}$.

Theorem 2. *Let $\mathcal{A}$ be an algebra with one generator element g and let $(\mathcal{M}, \rho)$ be a finite or countable infinite dimensional representation of $\mathcal{A}$. Let $(\mathcal{M}, \tilde{\rho})$ be a perturbed version of $(\mathcal{M}, \rho)$ associated to the perturbation model in eqn. (19). If $p_{\mathcal{A}} \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$, then*

$$\|D_p \mathbf{T}(\mathbf{S})\| \leq (1+\delta) \left(L_0 \sup_{\mathbf{S}} \|\mathbf{T}(\mathbf{S})\| + L_1 \sup_{\mathbf{S}} \|D_{\mathbf{T}}(\mathbf{S})\| \right) \tag{26}$$

Proof. See Section VII-B □

It is worth pointing out that the constants involved in the upper bound of eqn. (26) depend on the properties of the filters and the difference between the eigenvectors of $\mathbf{S}$ and $\mathbf{T}_r$. Therefore, the difference between the eigenvectors of these operators do not determine if $p(\mathbf{S})$ is stable or not, although the absolute value of the stability constants increase proportionally to δ.

From theorems 1 and 2 we can state the notion of stability for algebraic filters in the following corollary.

Corollary 1. *Let $\mathcal{A}$ be an algebra with one generator element g and let $(\mathcal{M}, \rho)$ be a finite or countable infinite dimensional representation of $\mathcal{A}$. Let $(\mathcal{M}, \tilde{\rho})$ be a perturbed version of $(\mathcal{M}, \rho)$ related by the perturbation model in eqn. (19). Then, if $p_{\mathcal{A}} \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ the operator $p(\mathbf{S})$ is stable in the sense of definition 5 with $C_0 = (1+\delta)L_0$ and $C_1 = (1+\delta)L_1$.*

Proof. Replace (26) from Theorem 2 into (25) from Theorem 1 and reorder terms. □

B. Stability of Algebraic Filters in Algebras with multiple geneators

The stability results presented in previous subsection can be extended naturally to operators associated to representations of algebras with multiple generators. Before doing so, we extend the notions used to characterize subsets of algebras. If $\mathcal{A}$ is the algebra generated by $\{g_i\}_{i=1}^{m}$ and $g_i g_j = g_j g_i$ for all i, j with coefficients in $\mathbb{R}$, we use the isomorphism $\iota : \mathcal{A} \to \mathcal{A}_{\mathbb{C}}$ given by $\iota(g_i) = \lambda_i \in \mathbb{C}$. Then, we can use $\iota\left(p_{\mathcal{A}}(g_1, \ldots, g_m)\right) = p(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) \in \mathbb{C}^m$ to characterize elements in $\mathcal{A}$.

Therefore, stating conditions on $p(\boldsymbol{\lambda})$ we can define subsets of $\mathcal{A}$ via ι^{-1}. In the following definition we extend the concepts presented in previous subsection.

Definition 7. *Let $p : \mathbb{C}^m \to \mathbb{C}$ be a multivariate function. Then, it is said that p is Lipschitz if there exists $L_0 > 0$ such that*

$$|p(\boldsymbol{\lambda}) - p(\boldsymbol{\mu})| \leq L_0 \|\boldsymbol{\lambda} - \boldsymbol{\mu}\| \tag{27}$$

for all $\boldsymbol{\lambda}, \boldsymbol{\mu} \in \mathbb{C}^m$. Additionally, it is said that $p(\boldsymbol{\lambda})$ is Lipschitz integral if there exists $L_1 > 0$ such that

$$\left| \lambda_i \frac{\partial p(\boldsymbol{\lambda})}{\partial \lambda_i} \right| \leq L_1 \quad \text{for all } i \in \{1, \dots m\}, \tag{28}$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$.

With this notion at hand, we are ready to extend the stability theorems.

Theorem 3. *Let $\mathcal{A}$ be an algebra generated by $\{g_i\}_{i=1}^m$ and let $(\mathcal{M}, \rho)$ be a representation of $\mathcal{A}$ with $\rho(g_i) = \mathbf{S}_i \in End(\mathcal{M})$ for all i. Let $\tilde{\rho}(g_i) = \tilde{\mathbf{S}}_i \in End(\mathcal{M})$ where the pair $(\mathcal{M}, \tilde{\rho})$ is a perturbed version of $(\mathcal{M}, \rho)$ and $\tilde{\mathbf{S}}_i$ is related to $\mathbf{S}_i$ by the perturbation model in eqn. (17). Then, for any $p_{\mathcal{A}} \in \mathcal{A}$ we have*

$$\left\| p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x} \right\| \leq \|\mathbf{x}\| \sum_{i=1}^{m} \left(\left\| D_{p|\mathbf{S}_i}(\mathbf{S})\mathbf{T}(\mathbf{S}_i) \right\| + \mathcal{O}\left(\|\mathbf{T}(\mathbf{S}_i)\|^2 \right) \right) \tag{29}$$

where $D_{p|\mathbf{S}_i}(\mathbf{S})$ is the partial Fréchet derivative of p on $\mathbf{S}_i$.

Proof. See Section VII-A □

Notice that in eqn. (29) we naturally add the contribution associated to each generator. Therefore, to guarantee stability we must have stability in each generator. Now, we show how the Fréchet derivative of $\mathbf{T}(\mathbf{S})$ is involved in the stability properties when considering multiple generators.

Theorem 4. *Let $\mathcal{A}$ be an algebra with m generators $\{g_i\}_{i=1}^m$ and $g_i g_j = g_j g_i$ for all $i, j \in \{1, \dots m\}$. Let $(\mathcal{M}, \rho)$ be a finite or countable infinite dimensional representation of $\mathcal{A}$ and $(\mathcal{M}, \tilde{\rho})$ a perturbed version of $(\mathcal{M}, \rho)$ related by the perturbation model in eqn. (19). Then, if $p_{\mathcal{A}} \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ it holds that*

$$\begin{aligned} &\left\| D_{p|\mathbf{S}_i}(\mathbf{S})\mathbf{T}(\mathbf{S}_i) \right\| \\ &\quad \leq (1+\delta) \left(L_0 \sup_{\mathbf{S}_i \in \mathcal{S}} \|\mathbf{T}(\mathbf{S}_i)\| + L_1 \sup_{\mathbf{S}_i \in \mathcal{S}} \|D_{\mathbf{T}}(\mathbf{S}_i)\| \right) \end{aligned} \tag{30}$$

Proof. See Section VII-B □

It is important to remark that the upper bound in eqn. (30) is defined by the largest perturbation in a given generator although the constants associated are determined completely by the properties of the filters.

From theorems 3 and 4 we can state the stability results for filters in algebras with multiple generators in the following corollary.

Corollary 2. *Let $\mathcal{A}$ be an algebra with generators $\{g_i\}_{i=1}^m$ and $g_i g_j = g_j g_i$ for all i, j. Let $(\mathcal{M}, \rho)$ be a finite or countable infinite dimensional representation of $\mathcal{A}$ and $(\mathcal{M}, \tilde{\rho})$ be a perturbed version of $(\mathcal{M}, \rho)$ related by the perturbation model in eqn. (19). Then, if $p_{\mathcal{A}} \in \mathcal{A}_{L_0} \cup \mathcal{A}_{L_1}$ the operator $p(\mathbf{S})$ is stable in the sense of definition 5 with $C_0 = m(1+\delta)L_0$ and $C_1 = m(1+\delta)L_1$.*

Proof. Replacing eqn. (30) from theorem 4 into eqn. (29) from theorem 3 and organazing the terms. □

C. Stability of Algebraic Neural Networks

The results in Theorems 1 to 4 and corollaries 1 and 2 can be extended to operators representing AlgNNs. We say that for a given AlgNN, $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^L$, a perturbed version of Ξ is given by $\tilde{\Xi} = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \tilde{\rho}_\ell)\}_{\ell=1}^L$ where $(\mathcal{A}_\ell, \mathcal{M}_\ell, \tilde{\rho}_\ell)$ is a perturbed version of $(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)$. For the sake of simplicity we present a theorem for AlgNNs with algebras with a single generator, but notice that these results can be easily stated for AlgNNs with multiple generators directly from theorems 3 and 4. To do so, we start highlighting in the following theorem the stability properties of the operators in the layer ℓ of an AlgNN.

Theorem 5. *Let $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^L$ be an algebraic neural network with L layers, one feature per layer and algebras $\mathcal{A}_\ell$ with a single generator. Let $\tilde{\Xi} = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \tilde{\rho}_\ell)\}_{\ell=1}^L$ be the perturbed version of Ξ by means of the perturbation model in eqn. (19). Then, if $\Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \mathcal{S}_\ell)$ and $\Phi\left(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \tilde{\mathcal{S}}_\ell\right)$ represent the mapping operators associated to Ξ and $\tilde{\Xi}$ in the layer ℓ respectively, we have*

$$\begin{aligned} &\left\| \Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \mathcal{S}_\ell) - \Phi\left(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \tilde{\mathcal{S}}_\ell\right) \right\| \leq \\ &C_\ell(1+\delta_\ell) \left(L_0^{(\ell)} \sup_{\mathbf{S}_\ell} \|\mathbf{T}^{(\ell)}(\mathbf{S}_\ell)\| + L_1^{(\ell)} \sup_{\mathbf{S}_\ell} \|D_{\mathbf{T}^{(\ell)}}(\mathbf{S}_\ell)\| \right) \|\mathbf{x}_{\ell-1}\| \end{aligned} \tag{31}$$

where C_ℓ is the Lipschitz constant of σ_ℓ, and $\mathcal{P}_\ell = \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ represents the domain of ρ_ℓ. The index (ℓ) makes reference to quantities and constants associated to the layer ℓ.

Proof. See Section VII-C1 □

This result, although simple, highlights the role of the maps σ_ℓ when perturbations are considered in each layer. In particular, we see that the effect of σ_ℓ is to scale $\boldsymbol{\Delta}_\ell$ by a constant but it does not change the nature or mathematical form of the perturbation. Notice also that σ_ℓ plays the role of a mixer that allows an AlgNN to provide selectivity without affecting the stability (see Section VIII).

Now we present in the following theorem the stability result for a general AlgNN with commutative algebras.

Theorem 6. *Let $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^L$ be an algebraic neural network with L layers, one feature per layer and algebras $\mathcal{A}_\ell$ with a single generator. Let $\tilde{\Xi} = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \tilde{\rho}_\ell)\}_{\ell=1}^L$ be the perturbed version of Ξ by means of the perturbation model in eqn. (19). Then, if $\Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L\right)$ and $\Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L\right)$ represent the mapping operators associated to Ξ and $\tilde{\Xi}$ respectively, we have*

$$\begin{aligned} &\left\| \Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L\right) - \Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L\right) \right\| \\ &\quad \leq \sum_{\ell=1}^{L} \boldsymbol{\Delta}_\ell \left(\prod_{r=\ell}^{L} C_r \right) \left(\prod_{r=\ell+1}^{L} B_r \right) \left(\prod_{r=1}^{\ell-1} C_r B_r \right) \|\mathbf{x}\| \end{aligned} \tag{32}$$

where C_ℓ is the Lipschitz constant of σ_ℓ, $\|\rho_\ell(\xi)\| \leq B_\ell$ for all $\xi \in \mathcal{P}_\ell$*, and $\mathcal{P}_\ell = \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ represents the domain of ρ_ℓ. The functions $\boldsymbol{\Delta}_\ell$ are given by*

$$\boldsymbol{\Delta}_\ell = (1+\delta_\ell) \left(L_0^{(\ell)} \sup_{\mathbf{S}_\ell} \|\mathbf{T}^{(\ell)}(\mathbf{S}_\ell)\| + L_1^{(\ell)} \sup_{\mathbf{S}_\ell} \|D_{\mathbf{T}^{(\ell)}}(\mathbf{S}_\ell)\| \right) \tag{33}$$

with the index (ℓ) indicating quantities and constants associated to the layer ℓ.

Proof. See Section VII-C2 □

Theorem 6 states how an AlgNN can be made stable by the selection of an appropriate subset of filters in the algebra, for a given perturbation model. It is worth pointing out that conditions like the ones obtained in [6] for GNNs can be considered particular instantiations of the conditions in Theorem 6. Additionally, notice that Theorem 6 can be easily extended to consider several features per layer, the reader can check the details of the proof of the theorem in Section VII-C2 where the analysis is performed considering multiple features.

Remark 2. It is important to highlight the fact that the perturbation model in eqn. (19) is *smooth* in the space of admissible $\mathbf{S}$ and this smoothness allows a consistent calculation of the Fréchet derivative of $\mathbf{T}(\mathbf{S})$. This can be considered as a consequence of the fact that deformations between arbitrary spaces can be measured according to the topology of the space. In particular, if a diffeomorphism is used to produce deformation in the domain of the signals of interest, it is possible to find an equivalent associated map that produces deformation of the set of operators acting on the signal. If notions of differentiability are used to measure the size of the original diffeomorphism, it is natural to find similar notions involved on the map acting on the operators, but with the difference that the differentiability is measured according to the topology of the new space.

VI. Spectral Operators

Before we proof the theorems presented in previous section, it is important to discuss the concept of *Fourier decomposition* in the context of representation theory of algebras. We want to highlight specially the role of the filters when a representation is compared to its perturbed version. In particular, we will show that there are essentially two factors that can cause differences between operators and their perturbed versions, the eigenvalues[1] and eigenvectors. Additionally, we show how the algebra can only affect one of those sources. This is consistent with the fact that differences in the eigenvectors of the operators only affect the constants that are associated to the stability bounds.

We highlight that the concept of spectral decomposition exploited extensively in traditional signal processing can be considered as a particular case of the concept of direct sums[2] of *irreducible* representations. In this section we provide some basic definitions from the representation theory of algebras and its connection with the notion of spectral decomposition of signals in algebraic neural networks. This will be useful and necessary in Section VII as the norm of the operators considered can be calculated in the terms of an spectral representation. We start with the following definition.

Definition 8. *Let $(\mathcal{M},\rho)$ be a representation of $\mathcal{A}$. Then, a representation $(\mathcal{U},\rho)$ of $\mathcal{A}$ is a **subrepresentation** of $(\mathcal{M},\rho)$ if $\mathcal{U}\subseteq\mathcal{M}$ and $\mathcal{U}$ is invariant under all operators $\rho(a)$* for all $a\in\mathcal{A}$, *i.e. $\rho(a)u\in\mathcal{U}$ for all $u\in\mathcal{U}$ and $a\in\mathcal{A}$. A representation $(\mathcal{M}\neq 0,\rho)$ is **irreducible** or simple if the only subrepresentations of $(\mathcal{M}\neq 0,\rho)$ are $(0,\rho)$ and $(\mathcal{M},\rho)$.*

[1]As we will show later, this is indeed a particular case of a general notion of homomorphism between the algebra and an irreducible subrepresentation of $\mathcal{A}$.

[2]When considering representations of non-compact groups this concept is insufficient and it is replaced for the more general notion of *direct integral of Hilbert spaces* [22], [23]

The class of irreducible representations of an algebra $\mathcal{A}$ is denoted by $\text{Irr}\{\mathcal{A}\}$. Notice that the zero vector space and $\mathcal{M}$ induce themselves subrepresentations of $(\mathcal{M},\rho)$. In order to state a comparison between representations the concept of *homomorphism between representations* is introduced in the following definition.

Definition 9. *Let $(\mathcal{M}_1,\rho_1)$ and $(\mathcal{M}_2,\rho_2)$ be two representations of an algebra $\mathcal{A}$. A homomorphism or **interwining operator** $\phi:\mathcal{M}_1\to\mathcal{M}_2$ is a linear operator which commutes with the action of $\mathcal{A}$, i.e.*

$$\phi(\rho_1(a)v)=\rho_2(a)\phi(v). \tag{34}$$

A homomorphism ϕ is said to be an isomorphism of representations if it is an isomorphism of vectors spaces.

Notice from definition 9 a substantial difference between the concepts of isomorphism of vector spaces and isomorphism of representations. In the first case we can consider that two arbitray vector spaces of the same dimension (finite) are isomorphic, while for representations that condition is required but still the condition in eqn. (34) must be satisfied. For instance, as pointed out in [24] all the irreducible 1-dimensional representations of the polynomial algebra $\mathbb{C}[t]$ are non isomorphic.

As we have discussed before, the vector space $\mathcal{M}$ associated to $(\mathcal{M},\rho)$ provides the space where the signals are modeled. Therefore, it is of central interest to determine whether it is possible or not to *decompose* $\mathcal{M}$ in terms of simpler or smaller spaces consistent with the action of ρ. We remark that for any two representations $(\mathcal{M}_1,\rho_1)$ and $(\mathcal{M}_2,\rho_2)$ of an algebra $\mathcal{A}$, their direct sum is given by the representation $(\mathcal{M}_1\oplus\mathcal{M}_2,\rho)$ where $\rho(a)(\mathbf{x}_1\oplus\mathbf{x}_2)=(\rho_1(a)\mathbf{x}_1\oplus\rho_2(a)\mathbf{x}_2)$. We introduce the concept of indecomposability in the following definition.

Definition 10. *A nonzero representation $(\mathcal{M},\rho)$ of an algebra $\mathcal{A}$ is said to be **indecomposable** if it is not isomorphic to a direct sum of two nonzero representations.*

Indecomposable representations provide the *minimum units of information* that can be extracted from signals in a given space when the filters have a specific structure (defined by the algebra) [25]. The following theorem provides the basic building block for the decomposition of finite dimensional representations.

Theorem 7 (Krull-Schmit, [26])**.** *Any finite dimensional representation of an algebra can be decomposed into a finite direct sum of indecomposable subrepresentations and this decomposition is unique up to the order of the summands and up to isomorphism.*

The uniqueness in this result means that if $(\oplus_{r=1}^{r}V_i,\rho)\cong(\oplus_{j=1}^{s}W_j,\gamma)$ for indecomposable representations $(V_j,\rho_j),(W_j,\gamma_j)$, then $r=s$ and there is a permutation π of the indices such that $(V_i,\rho_i)\cong(W_{\pi(j)},\gamma_{\pi(j)})$ [26]. Although theorem 7 provides the guarantees for the decomposition of representation in terms of indecomposable representations, it is not applicable when infinite dimensional representations are considered. However, it is possible to overcome this obstacle taking into account that irreducible representations are indecomposable [24], [26], and they can be used then to build representations that are indecomposable. In particular, *irreducibility* plays a central role to decompose the invariance properties of the images of ρ on $\text{End}(\mathcal{M})$ [26]. Representations that allow a decomposition in terms of subrepresentations that are irreducible are called *completely reducible* and its formal description is presented in the following definition.

Definition 11 ([26]). *A representation $(\mathcal{M}, \rho)$ of the algebra $\mathcal{A}$ is said to be* ***completely reducible*** *if $(\mathcal{M}, \rho) = \bigoplus_{i \in I}(\mathcal{U}_i, \rho_i)$ with irreducible subrepresentations $(\mathcal{U}_i, \rho_i)$. The* ***length*** *of $(\mathcal{M}, \rho)$ is given by $|I|$.*

For a given $(\mathcal{U}, \rho_{\mathcal{U}}) \in \text{Irr}\{\mathcal{A}\}$ the sum of all irreducible subrepresentations of (V, ρ_V) that are equivalent (isomorphic) to $(\mathcal{U}, \rho_{\mathcal{U}})$ is represented by $V(\mathcal{U})$ and it is called the $\mathcal{U}$-homogeneous component of (V, ρ_V). This sum is a direct sum, therefore it has a length that is well defined and whose value is called the *multiplicity* of $(\mathcal{U}, \rho_{\mathcal{U}})$ and is represented by $m(\mathcal{U}, V)$ [26]. Additionally, the sum of all irreducible subrepresentations of (V, ρ_V) will be denoted as $\text{soc}\{V\}$. It is possible to see that a given representation (V, ρ_V) is completely reducible if and only if $(V, \rho_V) = \text{soc}\{S\}$ [26]. The connection between $\text{soc}\{V\}$ and $V(\mathcal{U})$ is given by the following proposition.

Proposition 1 (Proposition 1.31 [26]). *Let $(V, \rho_V) \in \mathsf{Rep}\{A\}$. Then $soc\{V\} = \bigoplus_{S \in Irr\{\mathcal{A}\}} V(S)$.*

Now, taking into account that any homogeneous component $V(\mathcal{U})$ is itself a direct sum we have that

$$\text{soc}\{V\} \cong \bigoplus_{S \in \text{Irr}\{\mathcal{A}\}} S^{\oplus m(\mathcal{U}, V)}. \tag{35}$$

Equation (35) provides the building block for the definition of Fourier decompositions in algebraic signal processing [9]. With all these concepts at hand we are ready to introduce the following definition.

Definition 12 (Fourier Decomposition). *For an algebraic signal model $(\mathcal{A}, \mathcal{M}, \rho)$ we say that there is a spectral or Fourier decomposition of $(\mathcal{M}, \rho)$ if*

$$(\mathcal{M}, \rho) \cong \bigoplus_{(\mathcal{U}_i, \phi_i) \in Irr\{\mathcal{A}\}} (\mathcal{U}_i, \phi_i)^{\oplus m(\mathcal{U}_i, \mathcal{M})} \tag{36}$$

where the $(\mathcal{U}_i, \phi_i)$ are irreducible subrepresentations of $(\mathcal{M}, \rho)$. Any signal $\mathbf{x} \in \mathcal{M}$ can be therefore represented by the map Δ given by

$$\begin{aligned} \Delta : \quad \mathcal{M} \to \bigoplus_{(\mathcal{U}_i, \phi_i) \in Irr\{\mathcal{A}\}} & (\mathcal{U}_i, \phi_i)^{\oplus m(\mathcal{U}_i, \mathcal{M})} \\ \mathbf{x} \mapsto \hat{\mathbf{x}} & \end{aligned} \tag{37}$$

known as the Fourier decomposition of $\mathbf{x}$ and the projection of $\hat{\mathbf{x}}$ in each $\mathcal{U}_i$ are the Fourier components represented by $\hat{\mathbf{x}}(i)$.

Notice that in eqn. (36) there are two sums, one dedicated to the non isomorphic subrepresentations (external) and another one (internal) dedicated to subrepresentations that are isomorphic. In this context, the sum for non isomorphic representations indicates the sum on the *frequencies* of the representation while the sum for isomorphic representations a sum of components associated to a given frequency. It is also worth pointing out that Δ is an intertwining operator, therefore, we have that $\Delta(\rho(a)\mathbf{x}) = \rho(a)\Delta(\mathbf{x})$. As pointed out in [8] this can be used to define a convolution operator as $\rho(a)\mathbf{x} = \Delta^{-1}(\rho(a)\Delta(\mathbf{x}))$. The projection of a filtered signal $\rho(a)\mathbf{x}$ on each $\mathcal{U}_i$ is given by $\phi_i(a)\hat{\mathbf{x}}(i)$ and the collection of all this projections is known as the *spectral representation* of the operator $\rho(a)$. Notice that $\phi_i(a)\hat{\mathbf{x}}(i)$ translates to different operations depending on the dimension of $\mathcal{U}_i$. For instance, if $\dim(\mathcal{U}_i) = 1$, $\hat{\mathbf{x}}(i)$ and $\phi_i(a)$ are scalars while if $\dim(\mathcal{U}_i) > 1$ and finite $\phi_i(a)\hat{\mathbf{x}}(i)$ is obtained as a matrix product.

Remark 3. The spectral representation of an operator indicated as $\phi_i(a)\hat{\mathbf{x}}(i)$ and eqns. (36) and (37) highlight one important fact that is essential for the discussion of the results in Section VII. For a completely reducible representation $(\mathcal{M}, \rho) \in \mathsf{Rep}\{\mathcal{A}\}$ the connection between the algebra $\mathcal{A}$ and the spectral representation is *exclusively* given by $\phi_i(a)$ which is acting on $\hat{\mathbf{x}}(i)$, therefore, it is not possible by the selection of elements or subsets of the algebra to do any modification on the spaces $\mathcal{U}_i$ associated to the irreducible components in eqn.(36). As a consequence, when measuring the similarities between two operators $\rho(a)$ and $\tilde{\rho}(a)$ associated to $(\mathcal{M}, \rho)$ and $(\mathcal{M}, \tilde{\rho})$, respectively, there will be two sources of error. One source of error that can be modified by the selection of $a \in \mathcal{A}$ and another one that will be associated with the differences between spaces $\mathcal{U}_i$ and $\tilde{\mathcal{U}}_i$, which are associated to the direct sum decomposition of $(\mathcal{M}, \rho)$ and $(\mathcal{M}, \tilde{\rho})$, respectively. This point was first elucidated in [6] for the particular case of GNNs, but it is part of a much more general statement that becomes more clear in the language of algebraic signal processing.

Example 13 (Discrete signal processing). In CNNs the filtering is defined by the polynomial algebra $\mathcal{A} = \mathbb{C}[t]/(t^N - 1)$, therefore, in a given layer the spectral representation of the filters is given by

$$\begin{aligned} \rho(a)\mathbf{x} &= \sum_{i=1}^{N} \phi_i \left(\sum_{k=0}^{K-1} h_k t^k \right) \hat{\mathbf{x}}(i)\mathbf{u}_i \\ &= \sum_{i=1}^{N} \sum_{k=0}^{K-1} h_k \phi_i(t)^k \hat{\mathbf{x}}(i)\mathbf{u}_i = \sum_{i=1}^{N} \sum_{k=0}^{K-1} h_k \left(e^{-\frac{2\pi i j}{N}} \right)^k \hat{\mathbf{x}}(i)\mathbf{u}_i, \end{aligned}$$

with $a = \sum_{k=0}^{K-1} h_k t^k$ and where the $\mathbf{u}_i(v) = \frac{1}{\sqrt{N}} e^{\frac{2\pi j v i}{N}}$ are the column vectors of the traditional DFT matrix, while $\phi_i(t) = e^{-\frac{2\pi j i}{N}}$ is the eigenvalue associated to $\mathbf{u}_i$. Here $\hat{\mathbf{x}}$ represents the DFT of $\mathbf{x}$.

Example 14 (Graph signal processing). Taking into account that the filtering in each layer of a GNN is defined by a polynomial algebra, the spectral representation of the filter is given by

$$\begin{aligned} \rho(a)\mathbf{x} &= \sum_{i=1}^{N} \phi_i \left(\sum_{k=0}^{K-1} h_k t^k \right) \hat{\mathbf{x}}(i)\mathbf{u}_i \\ &= \sum_{i=1}^{N} \sum_{k=0}^{K-1} h_k \phi_i(t)^k \hat{\mathbf{x}}(i)\mathbf{u}_i = \sum_{i=1}^{N} \sum_{k=0}^{K-1} h_k \lambda_i^k \hat{\mathbf{x}}(i)\mathbf{u}_i \end{aligned} \tag{38}$$

with $a = \sum_{k=0}^{K-1} h_k t^k$, and where the $\mathbf{u}_i$ are given by the eigenvector decomposition of $\rho(t) = \mathbf{S}$, where $\mathbf{S}$ could be the adjacency matrix or the Laplacian of the graph, while $\phi_i(t) = \lambda_i$ being λ_i the eigenvalue associated to $\mathbf{u}_i$. The projection of $\mathbf{x}$ in each subspace $\mathcal{U}_i$ is given by $\hat{\mathbf{x}}(i) = \langle \mathbf{u}_i, \mathbf{x} \rangle$, and if $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{S}$ we have the widely known representation $\hat{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$ [20].

Example 15 (Group signal processing). Considering the Fourier decomposition on general groups [14]–[16], we obtain the spectral representation of the algebraic filters as

$$\boldsymbol{a} * \mathbf{x} = \sum_{u,h \in G} \boldsymbol{a}(uh^{-1}) \sum_{i,j,k} \frac{d_k}{|G|} \hat{\mathbf{x}} \left(\boldsymbol{\varphi}^{(k)} \right)_{i,j} \boldsymbol{\varphi}_{i,j}^{(k)}(h) hu,$$

where $\hat{\mathbf{x}}(\boldsymbol{\varphi}^{(k)})$ represents the Fourier components associated to the kth irreducible representation with dimension d_k and $\boldsymbol{\varphi}^{(k)}$ is the associated unitary element. We can see that the kth element in this decomposition is $\sum_{i,j}\mathbf{x}(\varphi^{(k)})_{i,j}\sum_{u,h}\frac{d_k}{|G|}\boldsymbol{a}(uh^{-1})\boldsymbol{\varphi}_{i,j}^{(k)}(h)hu$.

Example 16 (Graphon signal processing)**.** According to the spectral theorem [27], [28], it is possible to represent the action of a compact normal operator S as $S\mathbf{x}=\sum_i\lambda_i\langle\boldsymbol{\varphi}_i,\mathbf{x}\rangle\boldsymbol{\varphi}_i$ where λ_i and $\boldsymbol{\varphi}_i$ are the eigenvalues and eigenvectors of S, respectively, and $\langle\cdot\rangle$ indicates an inner product. Then, the spectral representation of the filtering of a signal in the layer ℓ is given by

$$\rho_\ell\left(p(t)\right)\mathbf{x}=\sum_i p(\lambda_i)\langle\mathbf{x},\boldsymbol{\varphi}_i\rangle\boldsymbol{\varphi}_i=\sum_i\phi_i(p(t))\hat{\mathbf{x}}_i\boldsymbol{\varphi}_i,$$

where $\phi_i(p(t))=p(\lambda_i)$.

VII. Proof of Theorems

Let us start defining some notation. Let $\boldsymbol{\pi}(a_1\mathbf{A}_1,\ldots,a_r\mathbf{A}_r)$ the operator that represents the sum of all the products of the operators $\mathbf{A}_1,\ldots,\mathbf{A}_r$ that appear $a_1,a_2,\ldots,a_r$ respectively. For instance, $\boldsymbol{\pi}(2\mathbf{A},\mathbf{B})=\mathbf{AAB}+\mathbf{ABA}+\mathbf{BAA}$. Additionally, when considering all summation and product symbols the following convention is used $\sum_{i=a}^b F(i)=0$ if $b<a$, and $\prod_{i=a}^b F(i)=0$ if $b<a$.

A. Proof of Theorems 1 and 3

Proof. We say that $p(\mathbf{S})$ as a function of $\mathbf{S}$ is Fréchet differentiable at $\mathbf{S}$ if there exists a bounded linear operator $D_p:\text{End}(\mathcal{M})^m\to\text{End}(\mathcal{M})$ such that [29], [30]

$$\lim_{\|\boldsymbol{\xi}\|\to 0}\frac{\|p(\mathbf{S}+\boldsymbol{\xi})-p(\mathbf{S})-D_p(\mathbf{S})\left\{\boldsymbol{\xi}\right\}\|}{\|\boldsymbol{\xi}\|}=0 \quad (39)$$

which in Landau notation can be written as

$$p(\mathbf{S}+\boldsymbol{\xi})=p(\mathbf{S})+D_p(\mathbf{S})\left\{\boldsymbol{\xi}\right\}+o(\|\boldsymbol{\xi}\|). \quad (40)$$

Calculating the norm in eqn. (40) and applying the triangle inequality we have: $\|p(\mathbf{S}+\boldsymbol{\xi})-p(\mathbf{S})\|\leq\|D_p(\mathbf{S})\left\{\boldsymbol{\xi}\right\}\|+\mathcal{O}\left(\|\boldsymbol{\xi}\|^2\right)$ for all $\boldsymbol{\xi}=(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_m)\in\text{End}(\mathcal{M})^m$. Now, taking into account that (see [31] pages 69-70)

$$\|D_p(\mathbf{S})\left\{\boldsymbol{\xi}\right\}\|\leq\sum_{i=1}^m\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\boldsymbol{\xi}_i\right\}\right\| \quad (41)$$

we have

$$\|p(\mathbf{S}+\boldsymbol{\xi})-p(\mathbf{S})\|\leq\sum_{i=1}^m\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\boldsymbol{\xi}_i\right\}\right\|+\mathcal{O}\left(\|\boldsymbol{\xi}\|^2\right),$$

where $D_{p|\mathbf{S}_i}(\mathbf{S})$ is the partial Frechet derivative of $p(\mathbf{S})$ on $\mathbf{S}_i$. Then, taking into account that $\|p(\mathbf{S}+\boldsymbol{\xi})\mathbf{x}-p(\mathbf{S})\mathbf{x}\|\leq\|\mathbf{x}\|\,\|p(\mathbf{S}+\boldsymbol{\xi})-p(\mathbf{S})\|$ and selecting $\boldsymbol{\xi}_i=\mathbf{T}(\mathbf{S}_i)$ we complete the proof. □

B. Proof of Theorem 2

Proof. Taking into account the definition of the Fréchet derivative of p on $\mathbf{S}_i$ (see Appendix B) we have

$$\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{T}(\mathbf{S}_i)\right\}\right\|=\left\|\sum_{k_i=1}^{\infty}\mathbf{A}_{k_i}\boldsymbol{\pi}\left(\mathbf{T}(\mathbf{S}_i),(k_i-1)\mathbf{S}_i\right)\right\|,$$

and re-organizating terms we have

$$\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{T}(\mathbf{S}_i)\right\}\right\|=\left\|\sum_{\ell=1}^{\infty}\mathbf{S}_i^{\ell-1}\mathbf{T}(\mathbf{S}_i)\sum_{k_i=\ell}^{\infty}\mathbf{A}_{k_i}\mathbf{S}_i^{k_i-\ell}\right\|.$$

Taking into account eqn. (20), it follows that

$$\begin{aligned}\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{T}(\mathbf{S}_i)\right\}\right\|&=\\ \Bigg\|\sum_{\ell=1}^{\infty}&\left(\mathbf{T}_{0c,i}\mathbf{S}_i^{\ell-1}+\mathbf{S}_i^{\ell-1}\mathbf{P}_{0,i}\right)\sum_{k_i=\ell}^{\infty}\mathbf{A}_{k_i}\mathbf{S}_i^{k_i-\ell}\\ &+\sum_{\ell=1}^{\infty}\left(\mathbf{T}_{1c,i}\mathbf{S}_i^{\ell}+\mathbf{S}_i^{\ell-1}\mathbf{P}_{1,i}\mathbf{S}_i\right)\sum_{k=\ell}^{\infty}\mathbf{A}_{k_i}\mathbf{S}^{k_i-\ell}\Bigg\|.\end{aligned} \quad (42)$$

Applying the triangle inequality and distribuiting the sum we have

$$\begin{aligned}\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{T}(\mathbf{S}_i)\right\}\right\|&\leq\left\|\mathbf{T}_{0c,i}\sum_{\ell=1}^{\infty}\sum_{k_i=\ell}^{\infty}\mathbf{S}_i^{k_i-1}\mathbf{A}_{k_i}\right\|\\ &+\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{0,i}\right\}\right\|+\left\|\mathbf{T}_{1c,i}\sum_{\ell=1}^{\infty}\sum_{k_i=\ell}^{\infty}\mathbf{S}^{k_i}\mathbf{A}_{k_i}\right\|\\ &+\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{0,i}\mathbf{S}_i\right\}\right\|\end{aligned} \quad (43)$$

Now, we analyze term by term in eqn. (43). For the first term we take into account that $\sum_{\ell=1}^{\infty}\sum_{k_i=\ell}^{\infty}\mathbf{S}_i^{k_i-1}\mathbf{A}_{k_i}=\sum_{k_i=1}^{\infty}k_i\mathbf{A}_{k_i}\mathbf{S}_i^{k_i-1}$ and we apply the product norm property taking into account that the filters belong to $\mathcal{A}_{L_0}$, which leads to

$$\begin{aligned}&\left\|\mathbf{T}_{0c,i}\sum_{\ell=1}^{\infty}\sum_{k_i=\ell}^{\infty}\mathbf{S}_i^{k_i-1}\mathbf{A}_{k_i}\right\|\\ &\leq\|\mathbf{T}_{0c,i}\|\left\|\sum_{k_i=1}^{\infty}k_i\mathbf{A}_{k_i}\mathbf{S}_i^{k_i-1}\right\|\leq L_0\|\mathbf{T}_{0,i}\|.\end{aligned} \quad (44)$$

For the second term in eqn. (43) we take into account that (see [29] page 84, [30] page 158 and [32] page 386): $\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{0,i}\right\}\right\|\leq L_0\|\mathbf{P}_{0,i}\|$ if p is Gâteaux differentiable, which is always true because p is Fréchet differentiable and with the fact that $\|\mathbf{P}_{0,i}\|\leq\delta\|\mathbf{T}_{0,i}\|$, we have $\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{0,i}\right\}\right\|\leq L_0\delta\|\mathbf{T}_{0,i}\|$.

For the third term in eqn. (43), we take into account that $\sum_{\ell=1}^{\infty}\sum_{k_i=\ell}^{\infty}\mathbf{S}_i^{k_i}\mathbf{A}_{k_i}=\sum_{k_i=1}^{\infty}k_i\mathbf{A}_{k_i}\mathbf{S}_i^{k_i}$ and we apply the norm product property taking into account that the filters belong to $\mathcal{A}_{L_1}$, which leads to

$$\begin{aligned}&\left\|\mathbf{T}_{1c,i}\sum_{\ell=1}^{\infty}\sum_{k_i=\ell}^{\infty}\mathbf{S}_i^{k_i}\mathbf{A}_{k_i}\right\|\\ &\leq\|\mathbf{T}_{1c,i}\|\left\|\sum_{k_i=1}^{\infty}k_i\mathbf{A}_{k_i}\mathbf{S}_i^{k_i}\right\|\leq L_1\|\mathbf{T}_{1,i}\|.\end{aligned} \quad (45)$$

Finally, for the fourth term we use the notation $\tilde{D}(\mathbf{S})\left\{\mathbf{P}_{1,i}\right\}=D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{1,i}\mathbf{S}_i\right\}$. We start taking into account that (see [33] pages 61 and 331) the eigenvalues of the operator $\tilde{D}(\mathbf{S})$ represented as ζ_{pq} are given by

$$\zeta_{pq}=\left\{\begin{array}{lll}\frac{p(\lambda_p)-p(\lambda_q)}{\lambda_p-\lambda_q}\lambda_q & \text{if} & \lambda_p\neq\lambda_q\\ \lambda_p p^{'}(\lambda_p) & \text{if} & \lambda_p=\lambda_q\end{array}\right., \quad (46)$$

then, taking into account that the filters belong to $\mathcal{A}_{L_1}$ we have $\|\tilde{D}(\mathbf{S})\|\leq L_1$, therefore $\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{1,i}\mathbf{S}_i\right\}\right\|=\left\|\tilde{D}(\mathbf{S})\left\{\mathbf{P}_{1,i}\right\}\right\|\leq L_1\|\mathbf{P}_{1,i}\|$. Additionally, with $\|\mathbf{P}_{1,i}\|\leq\delta\|\mathbf{T}_{1,i}\|$ it follows that $\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{P}_{1,i}\mathbf{S}_i\right\}\right\|\leq L_1\delta\|\mathbf{T}_{1,i}\|$.

Putting all these results together into eqn. (43) we reach

$$\left\|D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\mathbf{T}(\mathbf{S}_i)\right\}\right\| \leq (1+\delta)L_0\|\mathbf{T}_{0,i}\| + (1+\delta)L_1\|\mathbf{T}_{1,i}\|$$
$$\leq (1+\delta)\left(L_0 \sup_{\mathbf{S}_i\in\mathcal{S}}\|\mathbf{T}(\mathbf{S}_i)\| + L_1 \sup_{\mathbf{S}_i\in\mathcal{S}}\|D_{\mathbf{T}}(\mathbf{S}_i)\|\right)$$

□

C. Proof of Theorems 5 and 6

1) Proof of Theorem 5:

Proof. Taking into account eqns. (10), and (11) and the fact that the maps σ_ℓ are Lipschitz with constant C_ℓ we have that

$$\left\|\sigma_\ell\left(p(\mathbf{S}_\ell)\mathbf{x}_{\ell-1}\right) - \sigma_\ell\left(p(\tilde{\mathbf{S}}_\ell)\mathbf{x}_{\ell-1}\right)\right\| \leq C_\ell\mathbf{\Delta}_\ell\|\mathbf{x}_{\ell-1}\|, \quad (47)$$

where $\mathbf{\Delta}_\ell = \|p(\mathbf{S}_\ell) - p(\tilde{\mathbf{S}}_\ell)\|$, and whose value is determined by theorems 1 and 2. □

2) Proof of Theorem 6:

Proof. Before starting the calculations let us introduce some notation. Let $\varphi_f(\ell, g) = \rho_\ell\left(\xi^{fg}\right)$ denote the image of the filter $\xi^{fg} \in \mathcal{A}_\ell$ that process the gth feature coming from the layer $\ell - 1$ and that is associated to fth feature in layer ℓ. As indicated before, σ_ℓ indicates the Lipschitz mapping from layer ℓ to layer $\ell + 1$. The term $\mathbf{x}_{\ell-1}^g$ indicates the gth feature in the layer $\ell - 1$. Then, we have that:

$$\left\|\mathbf{x}_\ell^f - \tilde{\mathbf{x}}_\ell^f\right\| \leq \left\|\sigma_{\ell-1}\sum_{g_{\ell-1}}\varphi_{g_\ell}(\ell-1, g_{\ell-1})\sigma_{\ell-2}\right.$$
$$\sum_{g_{\ell-2}}\varphi_{g_{\ell-1}}(\ell-2, g_{\ell-2})\cdots\sigma_1\sum_{g_1}\varphi_{g_2}\mathbf{x} -$$
$$\sigma_{\ell-1}\sum_{g_{\ell-1}}\tilde{\varphi}_{g_\ell}(\ell-1, g_{\ell-1})\sigma_{\ell-2}\sum_{g_{\ell-2}}\tilde{\varphi}_{g_{\ell-1}}(\ell-2, g_{\ell-2})$$
$$\left.\cdots\sigma_1\sum_{g_1}\tilde{\varphi}_{g_2}(1, g_1)\mathbf{x}\right\|. \quad (48)$$

In order to exapand eqn. (48) we start pointing out that:

$$A_{\ell+1}\sigma_\ell(a) - \tilde{A}_{\ell+1}\sigma_\ell(\tilde{a}) =$$
$$(A_{\ell+1} - \tilde{A}_{\ell+1})\sigma_\ell(a) + \tilde{A}_{\ell+1}(\sigma_\ell(a) - \sigma_\ell(\tilde{a})) \quad (49)$$

where $A_{\ell+1}$ and $\tilde{A}_{\ell+1}$ indicate filter operators and their perturbed versions, respectively. Now, noticing that $\|\sigma_\ell(a) - \sigma_\ell(b)\| \leq C_\ell\|a - b\|$, $\|A_{\ell+1} - \tilde{A}_{\ell+1}\| \leq \mathbf{\Delta}_\ell$ and $\|A_{\ell+1}\| \leq B_{\ell+1}$ we have the following relations

$$\sum_{g_k}\|\alpha - \tilde{\alpha}\| \leq \sum_{g_k}\left(\mathbf{\Delta}_k\|\sigma_{k-1}(\alpha)\| + B_kC_{k-1}\|\beta - \tilde{\beta}\|\right) \quad (50)$$

$$\sum_{g_k}\|\beta - \tilde{\beta}\| \leq \sum_{g_k}\sum_{g_{k-1}}\|\alpha - \tilde{\alpha}\| \quad (51)$$

$$\sum_{g_k}\|\sigma_{k-1}(\alpha)\| \leq \left(\prod_{r=1}^{k}F_r\right)\left(\prod_{r=1}^{k-1}C_rB_r\right)\|\mathbf{x}\| \quad (52)$$

where α and $\tilde{\alpha}$ represent sequences of symbols in eqn. (48) that start with a symbol of the type φ, while β and $\tilde{\beta}$ indicate a sequence of symbols that start with a summation symbol, and the tilde makes reference to symbols that are associated to the perturbed representations. The term $\mathbf{\Delta}_\ell$ is associated to the difference between the operators and their perturbed versions (see definition 5) in the layer ℓ and whose values are given in Theorems 1 and 2. Combining eqns. (50), (51) and (52) we have:

$$\left\|\mathbf{x}_L^f - \tilde{\mathbf{x}}_L^f\right\| \leq \sum_{\ell=1}^{L}\mathbf{\Delta}_\ell\left(\prod_{r=\ell}^{L}C_r\right)\left(\prod_{r=\ell+1}^{L}B_r\right)$$
$$\left(\prod_{r=\ell}^{L-1}F_r\right)\left(\prod_{r=1}^{\ell-1}C_rF_rB_r\right)\|\mathbf{x}\|, \quad (53)$$

where the products $\prod_{r=a}^{b}F(r) = 0$ if $b < a$. Now taking into account that

$$\left\|\Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L\right) - \Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L\right)\right\|^2 = \sum_{f=1}^{F_L}\left\|\mathbf{x}_L^f - \tilde{\mathbf{x}}_L^f\right\|^2$$

we have

$$\left\|\Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L\right) - \Phi\left(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L\right)\right\|$$
$$\leq \sqrt{F_L}\sum_{\ell=1}^{L}\mathbf{\Delta}_\ell\left(\prod_{r=\ell}^{L}C_r\right)\left(\prod_{r=\ell+1}^{L}B_r\right)\left(\prod_{r=\ell}^{L-1}F_r\right)$$
$$\left(\prod_{r=1}^{\ell-1}C_rF_rB_r\right)\|\mathbf{x}\| \quad (54)$$

□

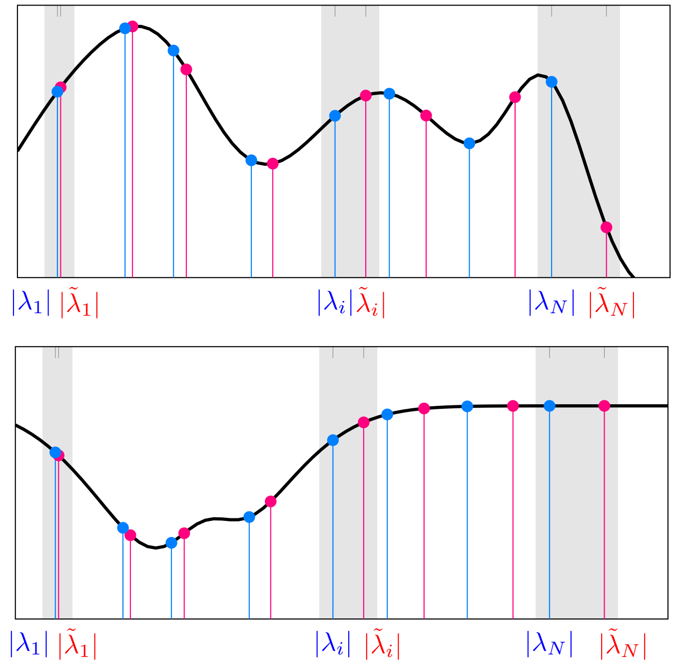

Figure 2. Filter properties and stability for algebraic operators considering algebras with a single generator. (Top) We depict a Lipschitz filter where it is possible to see that an arbitrary degree of selectivity can be achieved in any part of the spectrum. (bottom) We depict a Lipschitz integral filter where we can see how the magnitude of the filters tends to a constant value as the size of $|\lambda_i|$ grows. As a consequence there is no discriminability in one portion of the spectrum.

VIII. Discussion

The mathematical form of the notion of stability introduced in definition (5), eqn. (22) is uncannily similar to the expressions associated to the stability conditions stated in [1], [2] when the perturbation operator τ considered was affecting directly the domain of the signals. This is consistent with the fact that the size of the perturbation on the operators is the size of an induced diffeomorphism $\mathbf{T}$ acting on End$(\mathcal{M})$. Measuring the size of perturbations in this way, although less intuitive, provides an alternative way to handle and interpret perturbations on irregular domains.

The nature and severity of the perturbations, imposes restrictions on the behavior of the filters needed to guarantee stability. The more complex and severe the perturbation is the more conditions on the filters are necessary to guarantee stability. This in particular has implications regarding to the selectivity of the filters in some specific frequency bands. The trade-off between stability and selectivity in the filters of the AlgNN can be measured by the size of the upper bounds in Theorems 1 up to 6, in particular, the size of L_0 and L_1 associated to the boundedness of the derivatives of the elements in $\mathcal{A}_{L0}$ and $\mathcal{A}_{L1}$. The smaller the value of L_0, L_1 the more stable the operators but the less selectivity we have. In Fig. (2) the properties in frequency of Lipschtiz and Lipschitz integral filters are depicted, where it is possible to see how the selectivity on portions of the spectrum is affected by properties that at the same time provide stability conditions for the perturbation models considered.

It is important to remark that the function $\sigma_\ell = P_\ell \circ \eta_\ell$ composed by the projection operator P_ℓ and the nonlinearity function η_ℓ relocates information from one layer to the other performing a mapping between different portions of the spectrum associated to each of the spaces $\mathcal{M}_\ell$. As η_ℓ maps elements of $\mathcal{M}_\ell$ onto itself, we can see in light of the decomposition of $\mathcal{M}_\ell$ in terms of irreducible representations that η_ℓ is nothing but a relocator of information from one portion of the spectrum to the other. Additionally, the simplicity of η_ℓ provides a rich variety of choices that can be explored in future research.

The notion of differentiability between metric spaces or Banach spaces can be considered also using the notion of *Gâteaux derivative* which is considered a weak notion of differentiability. Although Gateaux differentiability is in general different from Fréchet differentiability, it is possible to show that when $\dim(\text{End}(\mathcal{M})) < \infty$ both notions are equivalent for Lipschtiz functions, but substantial differences may exist if $\dim(\text{End}(\mathcal{M})) = \infty$ even if the functions are Lipschitz [29], [30].

IX. Conclusions

We considered algebraic neural networks (AlgNN) with commutative algebras as a tool to unify convolutional architectures like CNNs and GNNs, synthesizing the algebraic structure by exploiting results from the representation theory of algebras and algebraic signal processing. Within this framework, we showed that AlgNNs can, in general, be stable to different types of perturbations, and the conditions under which the AlgNN operators are stable are determined by subsets of the algebra. We pointed out that the perturbations of the domain of the signals can be equivalently modeled as a perturbation of the representation or the signal model, and the degree of this perturbation can be measured by means of the Fréchet derivative of two functions, the image of the homomorhisms in End$(\mathcal{M})$ and the perturbation model $\mathbf{T}(\mathbf{S})$. The perturbation model considered provides enough expressive power to represent a wide variety of perturbations affecting the domain of the signals or the operator themselves directly. In particular, when considering the algebraic model for GNNs, the absolute and the relative perturbation models can be considered particular cases of the perturbation model considered in this work.

Appendix A
Perturbation Model

Proposition 2. *Let $\mathbf{T}_r, \mathbf{S} \in$ End$(\mathcal{M})$ two compact normal operators acting on $\mathcal{M}$. Then, $\mathbf{S}\mathbf{T}_r = \mathbf{T}_{cr}\mathbf{S} + \mathbf{S}\mathbf{P}_r$ with $\mathbf{T}_{cr} = \sum_i \mu_i \mathbf{u}_i \langle \mathbf{u}_i, \rangle$ and $\mathbf{P}_r \in$ End$(\mathcal{M})$. Additionally, if $\delta = \max_r \frac{\|\mathbf{P}_r\|}{\|\mathbf{T}_r\|}$ then*

$$\delta \leq \max_r \sum_{i,j} \left| \frac{\lambda_i}{\lambda_{max}} \right| \left\| \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{v}_j} - \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{u}_j} \right\|, \tag{55}$$

where $\mathbf{T}_{\mathbf{u}_i} = \mathbf{u}_i \langle \mathbf{u}_i, \cdot \rangle$, $\mathbf{T}_{\mathbf{v}_j} = \mathbf{v}_j \langle \mathbf{v}_j, \cdot \rangle$, $\mathbf{u}_i$ is the ith eigenvector of $\mathbf{S}$, $\mathbf{v}_j$ is the jth eigenvector of $\mathbf{P}_r$, $\langle \cdot, \cdot \rangle$ represents the inner product, λ_i is the ith eigenvalue of $\mathbf{S}$, μ_i is the ith eigenvalue of $\mathbf{T}_r$ and $\lambda_{max} = \sup_i\{|\lambda_i|\}$.

Proof. Let μ_i the ith eigenvalue of $\mathbf{T}_r$. Taking into account that $\|\mathbf{P}_r\|$ can be computed as $\|\mathbf{P}_r\| = \sup_{\mathbf{S}} \frac{\|\mathbf{S}\mathbf{P}_r\|}{\|\mathbf{S}\|}$, we have that

$$\begin{aligned}\frac{\|\mathbf{S}\mathbf{P}_r\|}{\|\mathbf{S}\|} &= \frac{\|\mathbf{S}\mathbf{T}_r - \mathbf{T}_{cr}\mathbf{S}\|}{\lambda_{max}} \\ &= \frac{1}{\lambda_{max}} \left\| \sum_{i,j} \lambda_i \mu_j \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{v}_j} - \sum_{i,j} \lambda_i \mu_j \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{u}_j} \right\| \end{aligned} \tag{56}$$

and using triangle inequality it follows that:

$$\begin{aligned}\frac{\|\mathbf{S}\mathbf{P}_r\|}{\|\mathbf{S}\|} &\leq \sum_{i,j} \left| \frac{\lambda_i}{\lambda_{max}} \right| |\mu_j| \left\| \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{v}_j} - \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{u}_j} \right\| \\ &\leq \|\mathbf{T}_r\| \sum_{i,j} \left| \frac{\lambda_i}{\lambda_{max}} \right| \left\| \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{v}_j} - \mathbf{T}_{\mathbf{u}_i}\mathbf{T}_{\mathbf{u}_j} \right\| \end{aligned} \tag{57}$$

□

Notice that the right hand side of eqn. (55) provides a measure of the differences between the eigenvectors of $\mathbf{S}$ and $\mathbf{T}_r$ that is weighted by the relative size of the eigenvalues λ_i. As expected, when the changes in the eigenvectors are associated to eigenvectors whose eigenvalues are small, value of δ is small.

Appendix B
Frechet Derivative $D_{p|\mathbf{S}_i}(\mathbf{S})$

First, notice that $p(\mathbf{S}) = \sum_{k_1,\ldots,k_m=0}^{\infty} h_{k_1 \ldots k_m} \mathbf{S}_1^{k_1} \ldots \mathbf{S}_m^{k_m} = \sum_{k_i=0}^{\infty} \mathbf{S}_i^{k_i} \mathbf{A}_{k_i}$, where $\mathbf{A}_{k_i} = \sum_{\substack{\{k_j\}=0 \\ j\neq i}}^{\infty} h_{k_1,\ldots,k_m} \prod_{\substack{j=1 \\ j\neq i}}^{m} \mathbf{S}_j^{k_j}$. Then, it follows that

$$p(\mathbf{S} + \boldsymbol{\xi}) - p(\mathbf{S}) = \sum_{k_i=0}^{\infty} (\mathbf{S}_i + \boldsymbol{\xi}_i)^{k_i} \mathbf{A}_{k_i} - \sum_{k_i=0}^{\infty} \mathbf{S}_i^{k_i} \mathbf{A}_{k_i} \tag{58}$$

for $\boldsymbol{\xi} = (\mathbf{0}, \ldots, \boldsymbol{\xi}_i, \ldots, \mathbf{0})$. Considering the expansion $(\mathbf{S}_i + \boldsymbol{\xi}_i)^{k_i} = \mathbf{S}_i^{k_i} + \boldsymbol{\xi}_i^{k_i} + \sum_{r=1}^{k-1} \boldsymbol{\pi}(r\mathbf{S}_i, (k_i - r)\boldsymbol{\xi}_i)$ for $k_i \geq 2$, eqn. (58) takes the form

$$\begin{aligned} &p(\mathbf{S} + \boldsymbol{\xi}) - p(\mathbf{S}) = \\ &\quad \sum_{k_i=1}^{\infty} \sum_{r=1}^{k_i-1} \boldsymbol{\pi}\left(r\boldsymbol{\xi}_i, (k_i - r)\mathbf{S}_i\right) \mathbf{A}_{k_i} + \sum_{k_i=1}^{\infty} \boldsymbol{\xi}_i^{k_i} \mathbf{A}_{k_i}. \end{aligned} \tag{59}$$

Separating the linear terms on $\boldsymbol{\xi}_i$ eqn. (59) leads to

$$p(\mathbf{S}+\boldsymbol{\xi}) - p(\mathbf{S}) = \sum_{k_i=1}^{\infty} \boldsymbol{\pi}\left(\boldsymbol{\xi}_i, (k_i-1)\mathbf{S}_i\right)\mathbf{A}_{k_i} + \sum_{k_i=2}^{\infty}\sum_{r=2}^{k_i-1} \boldsymbol{\pi}\left(r\boldsymbol{\xi}_i, (k_i-r)\mathbf{S}_i\right)\mathbf{A}_{k_i} + \sum_{k_i=2}^{\infty} \boldsymbol{\xi}^{k_i}\mathbf{A}_{k_i}. \quad (60)$$

Therefore, taking into account the definition of Fréchet derivative (see Section 1) it follows that

$$D_{p|\mathbf{S}_i}(\mathbf{S})\left\{\boldsymbol{\xi}_i\right\} = \sum_{k_i=1}^{\infty} \boldsymbol{\pi}\left(\boldsymbol{\xi}_i, (k_i-1)\mathbf{S}_i\right)\mathbf{A}_{k_i} \quad (61)$$

REFERENCES

[1] Stphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[2] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, Aug 2013.

[3] Joan Bruna Estrach, Arthur Szlam, and Yann LeCun. Learning stable group invariant representations with convolutional networks. 1 2013. 1st International Conference on Learning Representations, ICLR 2013 ; Conference date: 02-05-2013 Through 04-05-2013.

[4] Dongmian Zou and Gilad Lerman. Graph convolutional neural networks via scattering. *Applied and Computational Harmonic Analysis*, 1 2019.

[5] Fernando Gama, Joan Bruna Estrach, and Alejandro Ribeiro. Diffusion scattering transforms on graphs. 1 2019. 7th International Conference on Learning Representations, ICLR 2019 ; Conference date: 06-05-2019 Through 09-05-2019.

[6] Fernando Gama, Joan Bruna, and Alejandro Ribeiro. Stability properties of graph neural networks, 2019.

[7] Alberto Bietti and Julien Mairal. Invariance and stability of deep convolutional representations. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS17, page 62116221, Red Hook, NY, USA, 2017. Curran Associates Inc.

[8] Markus Pschel and Jos M. F. Moura. Algebraic signal processing theory, 2006.

[9] M. Puschel and J. M. F. Moura. Algebraic signal processing theory: Foundation and 1-d time. *IEEE Transactions on Signal Processing*, 56(8):3572–3585, Aug 2008.

[10] M. Puschel and J. M. F. Moura. Algebraic signal processing theory: 1-d space. *IEEE Transactions on Signal Processing*, 56(8):3586–3599, Aug 2008.

[11] J. Kovacevic and M. Puschel. Algebraic signal processing theory: Sampling for infinite and finite 1-d space. *IEEE Transactions on Signal Processing*, 58(1):242–257, Jan 2010.

[12] 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*, pages 6163–6166, 2013.

[13] A. Ortega, P. Frossard, J. Kovaevi, J. M. F. Moura, and P. Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828, May 2018.

[14] B. Steinberg. *Representation Theory of Finite Groups: An Introductory Approach*. Universitext. Springer New York, 2011.

[15] A. Terras. *Fourier Analysis on Finite Groups and Applications*. London Mathematical Society Student Texts. Cambridge University Press, 1999.

[16] W. Fulton, W.F.J. Harris, and J. Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York, 1991.

[17] L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society colloquium publications. American Mathematical Society, 2012.

[18] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra. Centrality measures for graphons: Accounting for uncertainty in networks. *IEEE Transactions on Network Science and Engineering*, 7(1):520–537, 2020.

[19] Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon signal processing, 2020.

[20] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro. Convolutional neural network architectures for signals supported on graphs. *IEEE Transactions on Signal Processing*, 67(4):1034–1049, 2019.

[21] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive computation and machine learning. MIT Press, 2016.

[22] G.B. Folland. *A Course in Abstract Harmonic Analysis*. Textbooks in Mathematics. CRC Press, 2016.

[23] A. Deitmar and S. Echterhoff. *Principles of Harmonic Analysis*. Universitext. Springer International Publishing, 2014.

[24] P.I. Etingof, O. Golberg, S. Hensel, T. Liu, A. Schwendner, D.V. , and E. Yudovina. *Introduction to Representation Theory*. Student mathematical library. American Mathematical Society, 2011.

[25] M. Barot. *Introduction to the Representation Theory of Algebras*. Springer International Publishing, 2014.

[26] M. Lorenz. *A Tour of Representation Theory*. Graduate studies in mathematics. American Mathematical Society, 2018.

[27] J.B. Conway. *A Course in Functional Analysis*. Graduate Texts in Mathematics. Springer New York, 1994.

[28] C.D. Aliprantis. *An Invitation to Operator Theory*. Graduate studies in mathematics. American Mathematical Society, 2002.

[29] Y. Benyamini and J. Lindenstrauss. *Geometric Nonlinear Functional Analysis*. Number v. 48, no. 1 in American Mathematical Society colloquium publications. American Mathematical Society, 2000.

[30] J. Lindenstrauss, D. Preiss, and J. Tišer. *Frechet Differentiability of Lipschitz Functions and Porous Sets in Banach Spaces*. Annals of Mathematics Studies. Princeton University Press, 2012.

[31] M.S. Berger. *Nonlinearity and Functional Analysis: Lectures on Nonlinear Problems in Mathematical Analysis*. Pure and Applied Mathematics. Elsevier Science, 1977.

[32] Joram Lindenstrauss, David Preiss, and Jaroslav Tivser. Fréchet differentiability of lipschitz functions via a variational principle. *Journal of the European Mathematical Society*, 12:385–412, 2010.

[33] N.J. Higham. *Functions of Matrices: Theory and Computation*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2008.