

# STABILITY OF ALGEBRAIC NEURAL NETWORKS TO SMALL PERTURBATIONS

Alejandro Parada-Mayorga and Alejandro Ribeiro

University of Pennsylvania

## ABSTRACT

Algebraic neural networks (AlgNNs) are composed of a cascade of layers each one associated to an algebraic signal model, and information is mapped between layers by means of a nonlinearity function. AlgNNs provide a generalization of neural network architectures where formal convolution operators are used, like for instance traditional neural networks (CNNs) and graph neural networks (GNNs). In this paper we study stability of AlgNNs on the framework of algebraic signal processing. We show how any architecture that uses a formal notion of convolution can be stable beyond particular choices of the shift operator, and this stability depends on the structure of subsets of the algebra involved in the model. We focus our attention on the case of algebras with a single generator.

**Index Terms**— Algebraic neural networks, algebraic signal processing, representation theory of algebras, representation theory, stability properties.

## 1. INTRODUCTION

Convolutional architectures play a central role on countless scenarios in machine learning and the numerical evidence that proves the advantages of using them is overwhelming [1, 2]. Additionally, theoretical insights like the ones developed in [3–5] and [6] have provided solid explanations about why such architectures work well. These analysis apparently different in nature, have been performed considering signals defined on different domains and with different notions of convolution but with remarkable similarities not only in the final results but also in how the derivations are performed, posing the question of whether there exists an explanation for this at a more structural level.

Several notions of stability have been considered widely in the literature, among which there are the first notion stability for traditional CNNs introduced in [3] by Mallat, and similar notions adapted for invariant scattering networks in [4] and networks affected by group actions in [5] on one side, while for irregular domains like GNNs notions of stability have been considered in [7, 8] and recently in [6] where concrete stability calculations are performed.

In this work we discuss the notion of stability of neural structures exploiting algebraic signal models whose particu-

lar instantiations lead to the convolutional operators used in traditional CNNs and GNNs. We formulate the definition of stability using the language of the representation theory of algebras providing concrete calculations of stability when algebras with a single generator are considered, and we point out that those results for CNNs and GNNs can be obtained as particular cases. Our results highlight the universality of stability beyond particular instantiations of algebraic models whenever the algebra involved is the same. This explains the notorious similarities between the analysis and final results about stability for CNNs and GNNs.

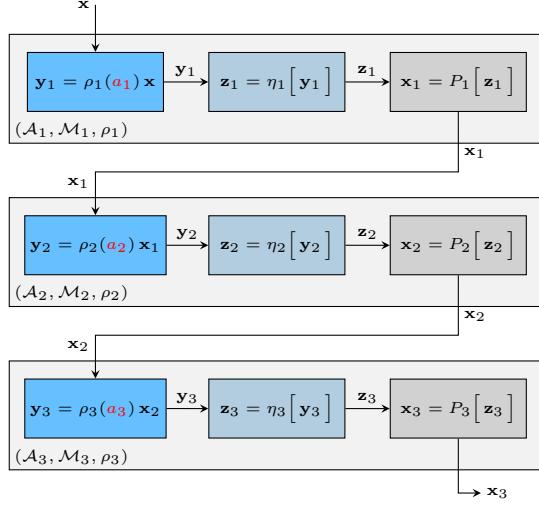
This paper is organized as follows. In Section 2 we discuss the basics about algebraic signal models and in Section 3 we introduce algebraic neural networks. In Section 4 we define the perturbations considered for the stability analysis and in Section 5 we discuss the stability results. Finally in section 6 we present some conclusions.

## 2. ALGEBRAIC SIGNAL MODELS

An algebraic signal model (ASM) can be defined by the triple  $(\mathcal{A}, \mathcal{M}, \rho)$ , where  $\mathcal{A}$  is an associative algebra with unity,  $\mathcal{M}$  is a vector space with inner product, and  $\rho : \mathcal{A} \rightarrow \text{End}(\mathcal{M})$  is a homomorphism between  $\mathcal{A}$  and the set of endomorphisms in  $\mathcal{M}$  [9–12]. In the context of representation theory of algebras, the pair  $(\mathcal{M}, \rho)$  is a representation of  $\mathcal{A}$ . Notice that  $\rho$  is a linear mapping that preserves the products in  $\mathcal{A}$ .

In the ASM the elements in  $\mathcal{M}$  are the signals, the elements in  $\mathcal{A}$  are the filters and  $\rho$  provides a physical realization associated to  $\mathcal{M}$  of the elements in  $\mathcal{A}$ . Then, the filtered version of a signal  $\mathbf{x} \in \mathcal{M}$  by an element in  $a \in \mathcal{A}$  is given by  $\mathbf{y} = \rho(a)\mathbf{x}$ . As pointed out in [9] the operation  $\rho(a)\mathbf{x}$  defines a general and formal notion of convolution between  $\rho(a)$  and  $\mathbf{x}$ , which is the notion used in this paper to study stability in algebraic neural networks.

Any element in an algebra  $\mathcal{A}$  can be expressed as a polynomial function of the *generators* of  $\mathcal{A}$ . For the discussion in this paper we focus on algebras that are generated by one element, i.e. all elements in  $\mathcal{A}$  can be expressed as  $a = p_{\mathcal{A}}(g)$  where  $g$  is the generator of  $\mathcal{A}$  and  $p_{\mathcal{A}}$  is a polynomial function. As a consequence of the properties of  $\rho$  as an homomorphism, we have that  $\rho(p_{\mathcal{A}}(g))$  is a polynomial  $p_{\mathcal{M}}$  expressed in terms of  $\rho(g)$ , therefore  $\rho(p_{\mathcal{A}}(g)) = p_{\mathcal{M}}(\rho(g))$ . Notice that the form of  $p_{\mathcal{A}}$  and  $p_{\mathcal{M}}$  is the same, therefore for the sake of sim-



**Fig. 1.** Example of an Algebraic Neural Network  $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^3$  with three layers. The input signal  $\mathbf{x}$  is mapped by  $\Xi$  into  $\mathbf{x}_3$ .

plicity of drop the subindex to denote  $\rho(p(g)) = p(\rho(g))$ . The operator  $\rho(g) = \mathbf{S}$  is called shift operator.

Particular cases of the algebraic models lead to signal processing frameworks well known in the literature. For instance using the polynomial algebra, which has a single generator, it is possible to derive discrete time signal processing (DTSP), graph signal processing (GSP) and graphon signal processing (WSP) considering different representations [9, 13], i.e. different choices of  $\mathcal{M}$  and  $\rho$ . DTSP uses a vector space (countably infinite dimensional) where the elements are sequences of square summable coefficients and the shift operator is the delay function, while in GSP the vector space is an  $N$ -dimensional vector space where  $N$  is the number of nodes in the graph and the shift operator could be the adjacency matrix of the graph while in WSP the vector space is the set of functions with finite energy in the interval  $[0, 1]$  and the shift is associated to an integral transform whose kernel is given by a graphon [9, 13].

### 3. ALGEBRAIC NEURAL NETWORKS

An algebraic neural network (AlgNN) consists of a stacked layered structure, where each layer is associated to an algebraic signal model. Then, if we have an AlgNN with  $L$  layers, there is an algebraic model  $(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)$  associated to the  $\ell$ th layer. The information is mapped from the layer  $\ell$  to the layer  $\ell+1$  by a map  $\sigma_\ell = P_\ell \circ \eta_\ell$  that is the composition of a pointwise nonlinearity and a pooling operator. For our discussion we consider  $\sigma_\ell$  as a Lipschitz map with zero as fixed point. In Fig. 1 a pictorial representation of an AlgNN is presented. Notice that the relationship between the output signal of the

layer  $\ell-1$  and the output signal of the layer  $\ell$  is then give by

$$\mathbf{x}_\ell = \sigma_\ell(\rho_\ell(a_\ell)\mathbf{x}_{\ell-1}) = \Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_{\ell-1}, \mathcal{S}_{\ell-1}), \quad (1)$$

with  $a_\ell \in \mathcal{A}_\ell$ . We use the symbol  $\Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_{\ell-1}, \mathcal{S}_{\ell-1})$  to emphasize that the filters used in the processing of the signal  $\mathbf{x}_{\ell-1}$  belong to the subset  $\mathcal{P}_\ell \subset \mathcal{A}_\ell$ . This will become relevant once we state stability results. The map associated to the whole AlgNN is represented by  $\Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L)$  acting on an input signal  $\mathbf{x}$ . We point out that the filtering in each layer of the AlgNN can be performed by means of several filter banks. This allows the generation of several features, case in which the output signal associated to the feature  $f$  in the layer  $\ell$  is given by  $\mathbf{x}_\ell^f$ .

### 4. DEFORMATIONS IN ALGEBRAIC SIGNAL MODELS

In previous sections we mentioned the role of  $\rho$  in the algebraic signal model  $(\mathcal{A}, \mathcal{M}, \rho)$  as a physical realization of the algebraic filters in  $\mathcal{A}$ , therefore it is natural in this context to associate the perturbation of the filters to  $\rho$ . One can think that  $\rho$  will provide an ideal realization of the filters in  $\mathcal{A}$  but that in practice one gets a realization  $\tilde{\rho}$  that is an approximate version of  $\rho$ . We consider deformations or perturbations of ASM analyzing changes in  $\rho$ . More specifically, we say that the triple  $(\mathcal{A}, \mathcal{M}, \tilde{\rho})$  is a perturbed version of  $(\mathcal{A}, \mathcal{M}, \rho)$  where  $\tilde{\rho}$  is defined according to

$$\tilde{\rho}(p(g)) = p(\tilde{\rho}(g)) = p(\tilde{\mathbf{S}}) \quad (2)$$

where  $\tilde{\mathbf{S}}$  is related to the shift operator  $\mathbf{S} = \rho(g)$  by means of  $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{T}(\mathbf{S})$ , and the term  $\mathbf{T}(\mathbf{S})$  is called the perturbation associated to  $(\mathcal{A}, \mathcal{M}, \tilde{\rho})$ . It is important to point out that  $\tilde{\rho}$  is not necessarily a homomorphism. As pointed out in [13] perturbations in the domain of the signals can be reframed in terms of this perturbation model.

For our discussion we focus on deformations of the type

$$\mathbf{T}(\mathbf{S}) = \epsilon \mathbf{I} + \mathbf{T}_1 \mathbf{S}, \quad (3)$$

which consists of an additive or absolute perturbation  $\epsilon \mathbf{I}$  where  $\epsilon > 0$  is a small scalar and  $\mathbf{I}$  is the identity matrix and a relative perturbation  $\mathbf{T}_1 \mathbf{S}$ . The operator  $\mathbf{T}_1$  is a compact normal operator with operator norm  $\|\mathbf{T}_1\| \ll 1$ . The measure of the non commutativity between the operator  $\mathbf{T}_1$  and  $\mathbf{S}$  is modeled as  $\mathbf{S} \mathbf{T}_1 = \mathbf{T}_{c1} \mathbf{S} + \mathbf{S} \mathbf{P}_1$ , and by the factor  $\delta$  given by

$$\delta = \max \frac{\|\mathbf{P}_1\|}{\|\mathbf{T}_1\|}, \quad (4)$$

where  $\mathbf{T}_{c1} = \sum_i \mu_i \mathbf{u}_i \langle \mathbf{u}_i, \cdot \rangle$ ,  $(\mu_i, \mathbf{u}_i)$  is the  $i$ th eigenpair  $\mathbf{T}_1$ ,  $\mathbf{u}_i$  are the eigenvectors of  $\mathbf{S}$ , and  $\langle \cdot, \cdot \rangle$  is the inner product. As proven in [13] the value of  $\delta$  is upper bounded by the weighted difference between the eigenvectors of  $\mathbf{x}$  and  $\mathbf{T}_1$ .

## 5. STABILITY RESULTS

Now we introduce the notion of stability for our discussion. We emphasize the notation presented in previous section, where  $(\mathcal{A}, \mathcal{M}, \tilde{\rho})$  denotes the perturbed version of  $(\mathcal{A}, \mathcal{M}, \rho)$ ,  $\rho(g) = \mathbf{S}$ ,  $\tilde{\rho}(g) = \tilde{\mathbf{S}}$  being  $g$  the generator of the algebra.

**Definition 1.** *Given  $p(\mathbf{S})$  and  $p(\tilde{\mathbf{S}})$  defined on the algebraic signal model  $(\mathcal{A}, \mathcal{M}, \rho)$  and  $(\mathcal{A}, \mathcal{M}, \tilde{\rho})$  we say the operator  $p(\mathbf{S})$  is Lipschitz stable if there exist constants  $C_0, C_1 > 0$  such that*

$$\begin{aligned} & \|p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x}\| \leq \\ & \left[ C_0 \sup_{\mathbf{S} \in \mathcal{S}} \|\mathbf{T}(\mathbf{S})\| + C_1 \sup_{\mathbf{S} \in \mathcal{S}} \|D_{\mathbf{T}}(\mathbf{S})\| + \mathcal{O}(\|\mathbf{T}(\mathbf{S})\|^2) \right] \|\mathbf{x}\|, \end{aligned} \quad (5)$$

for all  $\mathbf{x} \in \mathcal{M}$ . In (5)  $D_{\mathbf{T}}(\mathbf{S})$  is the Fréchet derivative of the perturbation operator  $\mathbf{T}$ .

We point out that the right hand side in eqn. (5) provides a measure of the size of the perturbation  $\mathbf{T}(\mathbf{S})$ . Then, the definition 1 says that the operators representing a given algebraic filter are stable to the action of a perturbation if the change in the operator is proportional to the size of the perturbation. It is worth pointing out that this is the same principle considered in the concept of stability presented in [3]. Additionally, as a particular case of this definition we obtain also the notion of stability used for GNNs in [6].

### 5.1. Characterizing subsets of the algebra

While determining if subsets of filters in the algebra are stable or not, a characterization of such subsets is necessary. For algebras with a single generator this can be done using functions of an independent variable. It is worth pointing out that when considering algebras like the polynomial algebra of a single variable, generated by one element, it is not necessary to assume that the independent variable of the algebra takes values in a particular field. However, this type of assumption can be useful to define subsets of the algebra. For our discussion we assume that the independent variable of the algebra takes values in  $\mathbb{C}$ .

In the the following definition, we describe the properties of the subsets of filters involved in the stability calculations.

**Definition 2.** *Let  $p : \mathbb{C} \rightarrow \mathbb{C}$  be a one variable function. Then, we say that  $p$  is Lipschitz if there exists  $L_0 > 0$  such that*

$$|p(\lambda) - p(\mu)| \leq L_0 |\lambda - \mu| \quad (6)$$

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

$$\left| \lambda \frac{dp(\lambda)}{d\lambda} \right| \leq L_1 \text{ for all } \lambda. \quad (7)$$

We denote by  $\mathcal{A}_{L_0}$  the subset of  $\mathcal{A}$  whose elements are Lipschitz functions with constant  $L_0$  and by  $\mathcal{A}_{L_1}$  the subset of elements in  $\mathcal{A}$  that are Lipschitz integral with constant  $L_1$ .

### 5.2. Stability of Algebraic Filters

Before stating concrete stability calculations for the particular type of deformation we are considering, we introduce a theorem, that provides essential insights for our discussion.

**Theorem 1.** *Let  $\mathcal{A}$  be an algebra generated by  $g$  and let  $(\mathcal{M}, \rho)$  be a representation of  $\mathcal{A}$  with  $\rho(g) = \mathbf{S} \in \text{End}(\mathcal{M})$ . Let  $\tilde{\rho}(g) = \tilde{\mathbf{S}} \in \text{End}(\mathcal{M})$  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. (3). Then, for any  $p \in \mathcal{A}$  we have*

$$\|p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x}\| \leq \|\mathbf{x}\| (\|D_p(\mathbf{S})\{\mathbf{T}(\mathbf{S})\}\| + \mathcal{O}(\|\mathbf{T}(\mathbf{S})\|^2)) \quad (8)$$

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

*Proof.* See [13] □

Theorem 1 show how the filters act directly on the deformation. In particular, we can see from the therm  $\|D_p(\mathbf{S})\{\mathbf{T}(\mathbf{S})\}\|$  that the Fréchet derivative of the physical representation of the filter acts directly on  $\mathbf{T}(\mathbf{S})$ , and this is independent from the properties of  $\mathbf{T}(\mathbf{S})$ , i.e. the perturbation is not restricted to be the one we are considering for our discussion.

Now we present a theorem where an upper bound of the right hand side of eqn. (8) is determined in terms of the properties of  $\mathbf{T}(\mathbf{S})$  for a specific type of filters.

**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. (3). If  $p_A \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ , then*

$$\|D_p(\mathbf{T}(\mathbf{S}))\| \leq (1+\delta) \left( L_0 \sup_{\mathbf{S}} \|\mathbf{T}(\mathbf{S})\| + L_1 \sup_{\mathbf{S}} \|D_{\mathbf{T}}(\mathbf{S})\| \right) \quad (9)$$

*Proof.* See [13] □

From theorem 2 and taking into account Theorem 1, we can conclude that AlgNNs are stable to the perturbations considered, indeed the norm of the Fréchet derivative of the filter acting on the perturbation is bounded by the size of the perturbation. Notice also that the non commutativity between the operators  $\mathbf{S}$  and  $\mathbf{T}_1$  does not change the functional form of the upper size in eqn. (9), however it does increase the size of the constants.

### 5.3. Stability of Algebraic Neural Networks

We have shown that algebraic filters defined in the algebraic context are stable to the deformations considered in the model (3). In this section we discuss how this results are inherited by AlgNN. This is associated directly to the fact that the functions  $\sigma_\ell$  mapping information between layers are Lipschitz functions. We start pointing out the effect of the functions  $\sigma_\ell$  mapping information between the layers of the AlgNN. Before doing so we clarify that the AlgNN  $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^L$  is perturbed into  $\tilde{\Xi} = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \tilde{\rho}_\ell)\}_{\ell=1}^L$  if  $(\mathcal{A}_\ell, \mathcal{M}_\ell, \tilde{\rho}_\ell)$  is a perturbed version of  $(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)$ .

**Theorem 3.** *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  $\Xi$  by means of the perturbation model in eqn. (3). Then, if  $\Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \mathcal{S}_\ell)$  and  $\Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \tilde{\mathcal{S}}_\ell)$  represent the mapping operators associated to  $\Xi$  and  $\tilde{\Xi}$  in the layer  $\ell$  respectively, we have*

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

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

*Proof.* See [13]  $\square$

From theorem 3 we see that the functions  $\sigma_\ell$  do not affect the stability properties of the algebraic filters although a modification of the constants is produced. This highlights the goodness of  $\sigma_\ell$  as they do not affect the stability and enrich the way the network handles information with different spectral content allowing an arbitrary degree of selectivity.

We introduce our final result of stability for a general AlgNN with  $L$  layers and one generator.

**Theorem 4.** *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  $\Xi$  by means of the perturbation model in eqn. (3). Then, if  $\Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L)$  and  $\Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L)$  represent the mapping operators associated to  $\Xi$  and  $\tilde{\Xi}$  respectively, we have*

$$\begin{aligned} \|\Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L) - \Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L)\| \leq \\ \sum_{\ell=1}^L \Delta_\ell \left( \prod_{r=\ell}^L C_r \right) \left( \prod_{r=\ell+1}^L B_r \right) \left( \prod_{r=1}^{\ell-1} C_r B_r \right) \|\mathbf{x}\| \quad (11) \end{aligned}$$

where  $C_\ell$  is the Lipschitz constant of  $\sigma_\ell$ ,  $\|\rho_\ell(\xi)\| \leq B_\ell$  for all  $\xi \in \mathcal{P}_\ell$ , and  $\mathcal{P}_\ell = \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$  is the domain of  $\rho_\ell$ . The functions  $\Delta_\ell$  are given by

$$\Delta_\ell = (1+\delta_\ell) \left( L_0^{(\ell)} \sup_{\mathbf{S}_\ell} \|\mathbf{T}^{(\ell)}(\mathbf{S}_\ell)\| + L_1^{(\ell)} \sup_{\mathbf{S}_\ell} \|D_{\mathbf{T}^{(\ell)}}(\mathbf{S}_\ell)\| \right) \quad (12)$$

with the index  $(\ell)$  indicating quantities and constants associated to the layer  $\ell$ .

*Proof.* See Section [13]  $\square$

This final result provides a guarantee of stability for AlgNNs at an algebraic level. It is worth remarking that:

- The representations used in each layer can be substantially different from one another, i.e. the pairs  $(\mathcal{M}_\ell, \rho_\ell)$  can change substantially from layer  $\ell$  to layer  $\ell+1$ , and still the stability results hold. This highlights the rich variety of options for the representation and processing of information with different types of operators without affecting the stability.
- A particular instantiation of the algebraic model considering the polynomial algebra and a vector space whose dimension equals the number of nodes in a graph, leads to results of stability for GNNs obtained in [6].
- The fact that AlgNNs inherit stability from the stability of the algebraic operators resides on the properties of the maps  $\sigma_\ell$  and the composition of several layers does not affect the functional expressions defining the stability but instead they increase the size of the constants.

We remark that this results are generalized for algebras with multiple generators, the reader interested in this extension can check the details of the proofs in [13] where the analysis is performed considering algebras with multiple generators.

## 6. CONCLUSIONS

We discussed algebraic neural networks (AlgNNs) considering algebras with a single generator in order to generalize the notion of stability when considering general convolution operators. Particular choices of the algebras and their representations in a AlgNN lead to traditional CNNs and GNNs, therefore the results obtained in this paper provide a unifying framework for the stability analysis. Indeed, we can say that stability transcend particular choices of spaces and operators whenever a formal representation is behind the structure of each layer. Additionally, we show that the physical realization of the algebraic filters as elements of  $\text{End}(\mathcal{M})$  acts directly on the perturbation by means of its Fréchet derivative, and it is precisely the action of this operator on the perturbation what determines the structure of the subsets of the algebra that will guarantee stability.

## 7. REFERENCES

- [1] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive computation and machine learning. MIT Press, 2016.
- [2] 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.
- [3] Stphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.
- [4] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, Aug 2013.
- [5] 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.
- [6] Fernando Gama, Joan Bruna, and Alejandro Ribeiro. Stability properties of graph neural networks, 2019.
- [7] Dongmian Zou and Gilad Lerman. Graph convolutional neural networks via scattering. *Applied and Computational Harmonic Analysis*, 1 2019.
- [8] 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.
- [9] Markus Puschel and Jos M. F. Moura. Algebraic signal processing theory, 2006.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Alejandro Parada-Mayorga and Alejandro Ribeiro. Algebraic neural networks: Stability to deformations, 2020.