

Algebraic Neural Networks: Stability to Deformations

Alejandro Parada-Mayorga  and Alejandro Ribeiro 

Abstract—We study algebraic neural networks (AlgNNs) with commutative algebras which unify diverse architectures such as Euclidean convolutional neural networks, graph neural networks, and group neural networks under the umbrella of algebraic signal processing. An AlgNN is a stacked layered information processing structure where each layer is conformed by an algebra, a vector space and a homomorphism between the algebra and the space of endomorphisms of the vector space. Signals are modeled as elements of the vector space and are processed by convolutional filters that are defined as the images of the elements of the algebra under the action of the homomorphism. We analyze stability of algebraic filters and AlgNNs to deformations of the homomorphism and derive conditions on filters that lead to Lipschitz stable operators. We conclude that stable algebraic filters have frequency responses – defined as eigenvalue domain representations – whose derivative is inversely proportional to the frequency – defined as eigenvalue magnitudes. It follows that for a given level of discriminability, AlgNNs are more stable than algebraic filters, thereby explaining their better empirical performance. This same phenomenon has been proven for Euclidean convolutional neural networks and graph neural networks. Our analysis shows that this is a deep algebraic property shared by a number of architectures.

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 overwhelming empirical evidence that shows the goodness of using convolutional neural networks (CNNs) and graph neural networks (GNNs) in machine learning raises interest in finding reasons that explain their performance. In this context, stability analyses of the operators representing the neural networks play a central role, with insights reported for both CNNs [1]–[4] and GNNs [5]–[7]. Although independent, these results are similar in form and nature. This fact raises the question of whether they descend from a common notion of stability and motivates the search for a framework where these results can be unified.

Manuscript received September 11, 2020; revised April 30, 2021; accepted May 24, 2021. Date of publication May 27, 2021; date of current version June 16, 2021. This work was supported by NSF-Simons MoDL, NSF-DMS-2031895 and Simons 814645. The associate editor coordinating the review of this manuscript and approving it for publication was Ms. Irène Waldspurger. (Corresponding author: Alejandro Parada-Mayorga.)

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6243 USA (e-mail: alejopm@seas.upenn.edu; aribeiro@seas.upenn.edu).

Digital Object Identifier 10.1109/TSP.2021.3084537

Stability of CNNs is rooted in the notion of Lipschitz-continuity to the action of diffeomorphisms introduced in [1] for the analysis of translation-invariant operators acting on $L^2(\mathbb{R}^n)$. Although initially derived for scattering transforms [1], [2] stability results are readily extendable to the analysis of convolutional neural networks [3], [4]. For GNNs the problem of formulating stability conditions has been considered in [5]–[7]. In [6] 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 [7] a related notion of stability is used to provide concrete results about the stability on GNNs.

However different, stability results for CNNs and GNNs have uncanny similarities. For instance, they both focus on signal perturbations that are modeled as deformations of the signal domain, they both analyze the effect of perturbations in the frequency (spectral) domain, and they both conclude that (graph or Euclidean) convolutional filters have instabilities associated with high frequency components (large eigenvalues). In our search for underlying common principles we adopt the formalism of *algebraic signal processing* (ASP) [8].

In general, signals are elements of a vector space $\mathcal{M}$ which we could process with any linear transformation in the algebra of endomorphisms of $\mathcal{M}$. In practice, learning is facilitated if we introduce a suitable class of convolutional filters to restrict the type of transformations that are allowable. In ASP, convolutional *algebraic filters* are defined as elements of a more restrictive algebra $\mathcal{A}$ that are mapped into the algebra of endomorphisms of $\mathcal{M}$ through a homomorphism ρ (Section II). In the case of signals supported on a graph with n nodes, the vector space is made up of vectors of length n and the space of endomorphisms is made up of square matrices of matching dimension. Choosing the algebra of polynomials of a single variable t and choosing a homomorphism that maps t to the Laplacian matrix of the graph results in graph convolutional filters expressed as polynomials of the graph's Laplacian. This is the usual definition of a graph filter [9]. In the case of signals in time the vector space is that of square summable sequences and convolutions can be written as polynomials on the time shift operator. This is the standard definition of convolution for discrete time signals [10, Ch. 2].

In this paper we leverage algebraic filters to introduce *algebraic neural networks* (AlgNNs) and study their stability

to deformations of the signal domain. In particular, the main contributions of this paper are:

- (C1) The definition of AlgNNs as layered information processing architectures in which individual layers are made up of algebraic convolutional filters (Section III).
- (C2) The introduction of perturbation models (Section IV) and stability properties (Section V) that are analogous to the notions of perturbation and stability considered in [1]–[7].
- (C3) The proof of stability theorems for algebraic filters (Theorems 1-4 and Corollaries 1 and 2) out of which stability theorems for AlgNNs follow (Theorems 5 and 6).

Our results are meaningful for algebras with a small number of generators (see Definition 2). This includes discrete time convolutions but does not include continuous time convolutions. Thus, we do not recover results in [1]–[4] as particular cases of our theorems. Rather our conclusions for discrete time CNNs are analogous to the conclusions that [1]–[4] reach for continuous time CNNs. This relatively minor technicality aside, our stability results for AlgNNs recover existing results for CNNs and GNNs (Section V-D). Our results also extend to other types of convolutional architectures like multidimensional CNNs – as used in image processing –, group neural networks and graphon neural networks. They also apply to as of yet unknown convolutional architectures. Indeed, the universality of stability properties is among the fundamental insights of this paper:

- (I1) The stability properties of convolutional filters and neural networks are universal.

This holds because the stability properties of convolutional architectures can be expressed in terms of the algebraic laws that govern the signal model in each layer as encoded in the algebra $\mathcal{A}$. To explain this statement we mention that representations of algebras admit spectral decompositions (Section VI). These decompositions permit the definition of Fourier transforms of signals and, more germane to our discussions, frequency representations of algebraic filters. These representations are defined as isomorphisms that map generators of the algebra $\mathcal{A}$ into scalar variables (Definitions 6 and 8). As such, frequency representations are functions with as many variables as generators as needed to generate the algebra. In cases of interest, this just means a function of a few variables, each of which we call a frequency. Remarkably, frequency representations depend on the choice of algebra but do not depend on the vector space where signals live. Ultimately, this is the reason why universal stability results are possible and it further leads to the following insights:

- (I2) Although perturbations are considered on filter operators, stability is determined by restrictions to certain subsets of the algebra. These restrictions are expressed in terms of filters' frequency representations (Section VII).
- (I3) Stability requires filter frequency responses that are flat for large values of the frequency variables. This limits the discriminability of algebraic filters (Section VIII).
- (I4) AlgNNs improve the stability vs discriminability tradeoff of algebraic filters because pointwise nonlinearities move

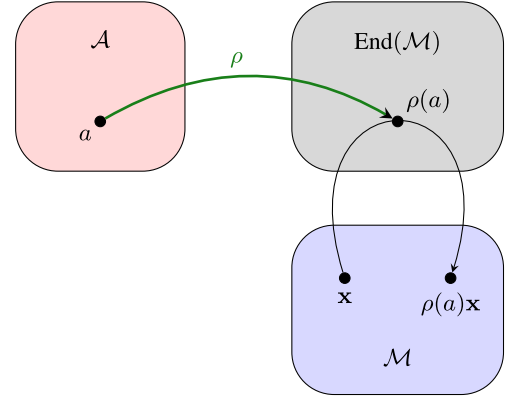


Fig. 1. Algebraic Signal Processing (ASP) Model. An ASP model is made up of a vector space $\mathcal{M}$ where signals $\mathbf{x}$ live and an Algebra $\mathcal{A}$ where filters a live. The homomorphism ρ ties the algebra and the vector space together by mapping the filter a to the linear function $\rho(a)$ in the space of endomorphisms of $\mathcal{M}$. The Algebra restricts the set of linear processing maps that can be applied to signals $\mathbf{x}$.

signal energy towards lower frequencies where signals can be better discriminated by filters with a given level of stability (Section VIII).

Insights (I3) and (I4) are the summary messages of this paper. We know from [1]–[4] that (I3) and (I4) explain the increased performance of CNNs relative to convolutional filters. We know from [5]–[7] that (I3) and (I4) explain the increased performance of GNNs relative to graph filters. As per (I1) we show here that the reason why these analogous properties hold is the shared algebraic structure of CNNs and GNNs. The universality of the result implies that (I3) and (I4) also explain performance improvements of CNNs with multidimensional inputs relative to multidimensional Euclidean convolutional filters, group neural networks relative to group filters, and graphon neural networks relative to graphon filters among any number of known and unknown convolutional information processing architectures. Our results are limited to commutative algebras with a small number of representers. Further work is needed to extend our results to these more general signal models (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]–[13]; see Figure 1. In ASP, a signal model is defined as the triple

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

in which $\mathcal{A}$ 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 the algebra $\mathcal{A}$ and the set of endomorphisms of the vector space $\mathcal{M}$. The elements in (1) are tied together by the notion of a representation which we formally define next.

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

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 $\mathcal{M}$ determines 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

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

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 that representations connect the algebra $\mathcal{A}$ to 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 $\mathcal{G} \subseteq \mathcal{A}$ 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 admissible 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 fact that ρ is a homomorphism that preserves operations implies that the filter's instantiation $\rho(a)$ can be written as

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

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 and for the rest of the paper we drop the subindices in the polynomials to simplify notation as it is generally understood from context to which set the independent variable of p belongs.

We restrict attention to commutative algebras $\mathcal{A}$. We also restrict the field $\mathbb{F}$ on which $\mathcal{M}$ and $\mathcal{A}$ are supported to be algebraically closed. If this doesn't hold our results apply to the corresponding algebraic extension. We point out that although not a formal requirement, our results are meaningful when the algebra $\mathcal{A}$ has a set of generators with a small number of elements.

We present examples to clarify ideas. Readers may skip ahead since they are not needed to understand the rest of the paper.

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 generated by $g = t$ with elements $a = \sum_{k=0}^{K-1} h_k t^k$. Consider the time shift operator S such that $S\mathbf{x}$ is the sequence with entries $(S\mathbf{x})_n = x_{n-1}$. Define the

homomorphism ρ in which the generator $g = t$ is mapped to $\rho(g) = \rho(t) = S$. Then, the filter $a = \sum_{k=0}^{K-1} h_k t^k$ is mapped to the endomorphism [cf.(3)]

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

Observe how the abstract polynomial $a = \sum_{k=0}^{K-1} h_k t^k$ is mapped to the polynomial $\rho(a) = \sum_{k=0}^{K-1} h_k S^k$. The latter is a concrete linear operator in the space of square summable sequences that we can use to process sequences $\mathbf{x}$ as per (2). This leads to the input output relationship

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

Since $(S\mathbf{x})_n = x_{n-1}$ it follows that $(S^k \mathbf{x})_n = x_{n-k}$ and that (5) represents a discrete time convolutional filter [10, Ch. 2].

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 $\mathbf{S} \in \mathbb{C}^{N \times N}$. We consider the homomorphism ρ in which the generator $g = t$ is mapped to the matrix representation $\mathbf{S}$ of the graph. Having chosen $\rho(t) = \mathbf{S}$ we use (3) to write

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

Analogously to (4), the abstract polynomial $\sum_{k=0}^{K-1} h_k t^k$ is mapped to the concrete polynomial $\sum_{k=0}^{K-1} h_k \mathbf{S}^k$. The latter is an $N \times N$ matrix that can be applied to signals $\mathbf{x}$ to produce outputs

$$\mathbf{y} = \left(\sum_{k=0}^{K-1} h_k \mathbf{S}^k\right) \mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}. \quad (7)$$

This is a representation of the graph convolutional filters used in graph signal processing (GSP) [14], [15]. Observe that (7) and (5) are similar but represent different operations. In (7) $\mathbf{x}$ is a vector and $\mathbf{S}^k$ a matrix power. In (5) $\mathbf{x}$ is a sequence and S^k is the composition of the time shift operator S . Their similarity arises from the common use of the algebra of polynomials. Their differences are because we use different vector spaces $\mathcal{M}$ and different homomorphisms ρ .

Example 3 (Discrete Signal Processing): We consider discrete time signals of length N with circular convolutions. To do that we consider the vector space $\mathcal{M} = \mathbb{C}^N$ and the algebra of polynomials modulo $t^N - 1$. I.e., filters $a \in \mathcal{A}$ are polynomials $a = \sum_{k=0}^{K-1} h_k t^k$ but we must have $K \leq N$ and monomial products use the rule $t^k = t^{k \bmod N}$. We consider the directed cyclic matrix $\mathbf{C}$ with exactly N nonzero entries $C_{n,m} = 1$ for $m = (n-1) \bmod N$. This matrix is such that $(\mathbf{C}\mathbf{x})_n = x_{(n-1) \bmod N}$. Using the homomorphism in which we map the generator $g = t$ to $\rho(g) = \rho(t) = \mathbf{C}$, filter instantiations

take the form

$$\rho \left(\sum_{k=0}^{K-1} h_k t^k \right) = \sum_{k=0}^{K-1} h_k \mathbf{C}^k. \quad (8)$$

The filter instantiation $\rho(a) = \sum_{k=0}^{K-1} h_k \mathbf{C}^k$ leads to the input output relationship

$$\mathbf{y} = \left(\sum_{k=0}^{K-1} h_k \mathbf{C}^k \right) \mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{C}^k \mathbf{x}. \quad (9)$$

Since $(\mathbf{C}\mathbf{x})_n = x_{(n-1) \bmod N}$ we have that $(\mathbf{C}^k \mathbf{x})_n = x_{(n-k) \bmod N}$. Thus, (9) is equivalent to the usual definition of circular convolutions [10, Ch. 8]. Observe that the homomorphism $\rho(a) = \sum_{k=0}^{K-1} h_k \mathbf{C}^k$ is indeed a homomorphism because the cyclic matrix $\mathbf{C}$ satisfies $\mathbf{C}^k = \mathbf{C}^{k \bmod N}$. This example illustrates that in some situations the choice of algebra and the choice of homomorphism are tied.

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 rich variety of signal processing frameworks. We highlight this richness with three more examples.

Example 4 (Image Processing): We represent images as square summable sequences with two indexes, $\mathbf{x} = \{x_{n,m}\}_{n,m \in \mathbb{Z}}$. We define the horizontal translation operator S_H such that $(S_H \mathbf{x})_{mn} = x_{m,n-1}$ and the vertical translation operator S_V such that $(S_V \mathbf{x})_{mn} = x_{m-1,n}$. Filters to process images are elements of the algebra of polynomials of two variables $a = \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} h_{k_1 k_2} t_1^{k_1} t_2^{k_2}$. This algebra has two generators $g_1 = t_1$ and $g_2 = t_2$ that we map to $\rho(t_1) = S_H$ and $\rho(t_2) = S_V$. This generator mapping defines the homomorphism ρ in which filters are mapped to instances

$$\rho \left(\sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} h_{k_1 k_2} t_1^{k_1} t_2^{k_2} \right) = \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} h_{k_1 k_2} S_H^{k_1} S_V^{k_2}. \quad (10)$$

The composed operator $S_H^{k_1} S_V^{k_2}$ applied to a sequence $\mathbf{x}$ translates horizontal and vertical indexes by k_1 and k_2 indexes. Thus, applying the operator in the right hand side of (10) to an image

$\mathbf{x}$ is equivalent to convolving the image with an 2-dimensional convolutional filter with coefficients $h_{k_1 k_2}$.

Example 5 (Signal Processing on Groups): Let $\mathcal{M} = \{\mathbf{x} : G \rightarrow \mathbb{C}\} = \{\sum_{g \in G} \mathbf{x}(g)g\}$ be the set of functions defined on the group G with values in $\mathbb{C}$ and $\mathcal{A} = \mathcal{M}$ the *group algebra*. The homomorphism is given by $\rho(a) = L_a$, with $L_a \mathbf{b} = \mathbf{a}\mathbf{b}$. Then, the action of ρ on elements of $\mathcal{M}$ is given by

$$\rho \left(\sum_{g \in G} \mathbf{a}(g)g \right) \mathbf{x} = \sum_{g \in G} \mathbf{a}(g)g\mathbf{x} = \sum_{g \in G} \sum_{h \in G} \mathbf{a}(g)\mathbf{x}(h)gh, \quad (11)$$

and making $u = gh$ we have that the filtering in (2) takes the form

$$\sum_{g,h \in G} \mathbf{a}(g)\mathbf{x}(h)gh = \sum_{u,h \in G} \mathbf{a}(uh^{-1})\mathbf{x}(h)u. \quad (12)$$

This is the standard representation of convolution of signals on groups [16]–[18]. We point out that (11) and (12) hold for any group but that not all group algebras are commutative. Results in Section V apply only when the group algebra is commutative.

Example 6 (Graphon Signal Processing): A graphon $W(u,v) : [0,1]^2 \rightarrow [0,1]$ is a bounded symmetric measurable function and graphon signals are square summable functions $\mathbf{x}(u) : [0,1] \rightarrow \mathbb{C}$. Graphons are intended to represent dense limits of graphs [19]–[22] and graphon signals dense limits of graph signals [21], [22]. To define graphon convolutional filters consider the algebra of polynomials of a single variable and define the graphon shift operator as

$$(\mathcal{W}\mathbf{x})(u) = \int_0^1 W(u,v)\mathbf{x}(v)dv, \quad (13)$$

Filters $a = \sum_{k=0}^{K-1} h_k t^k$ are mapped according to the homomorphism defined by the generator map $\rho(t) = \mathcal{W}$ resulting on filters that define the input-output relationship

$$\mathbf{y} = \rho \left(\sum_{k=0}^{K-1} h_k t^k \right) \mathbf{x} = \sum_{k=0}^{K-1} h_k \mathcal{W}^k \mathbf{x}. \quad (14)$$

This is the same definition of graphon convolutional filters introduced in [21] where they are shown to be limit objects of graph filters.

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 (5) leverage the shift invariance of time signals and the filters in (9) the cyclic invariance of periodic signals. The group convolutional filters in (12) generalize shift invariance with respect to an arbitrary group action. The graph convolutional filters in (7) engender signal processing that is independent of node labeling [23] and the graphon filters in Example 6 a generalization of this notion to dense domains [21]. 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.

Remark 1 (Shift Equivariance of Algebraic Filters): In restricting the linear transformations that can be applied to signals, the Algebra $\mathcal{A}$ reduces the complexity of the learning space. It is easier to learn coefficients of a filter than it is to learn entries of an arbitrary linear transform. In this statement, the *equivariance* of algebraic filters to applications of shift operators is important. Equivariance to applications of the shift operator means that applying a shift operator at the input of an algebraic filter is equivalent to applying the same shift operator at the output. Namely, that for all filters $a = p(\mathcal{G})$ and shift operators $\mathbf{S} \in \mathcal{S}$ we have

$$\mathbf{S}p(\mathcal{S})\mathbf{x} = p(\mathcal{S})\mathbf{S}\mathbf{x}. \quad (15)$$

This holds true for any commutative algebra. Equivariance to application of the shift operator is important in discrete time signal processing, discrete signal processing, image processing, and group signal processing. It implies that algebraic filters are equivariant to time shifts, cyclic shifts, translations, and actions of the group, respectively.

Remark 2 (Permutation Equivariance of Algebraic Filters): In learning with algebraic filters equivariance to permutations is also important. Equivariance to permutations means that a consistent permutation of the signal and the shift operator results in a consistent permutation of the output of the filter. Formally, let $\mathbf{P} \in \text{End}(\mathcal{M})$ be a permutation operator with adjoint $\mathbf{P}^T = \text{adj}(\mathbf{P})$. A permutation of the signal $\mathbf{x}$ is $\tilde{\mathbf{x}} = \mathbf{P}^T \mathbf{x}$ and a consistent permutation of the shift operator $\mathbf{S}$ is the endomorphism $\tilde{\mathbf{S}} = \mathbf{P}^T \mathbf{S} \mathbf{P}$. If we let $\tilde{\mathcal{S}}$ denote the set of permuted shift operators we must have,

$$p(\tilde{\mathcal{S}}) \tilde{\mathbf{x}} = \mathbf{P}^T (p(\mathcal{S})\mathbf{x}). \quad (16)$$

I.e., the output of processing a permuted signal $\tilde{\mathbf{x}}$ with the filter instantiated on the set of permuted shift operators $\tilde{\mathcal{S}}$ is equivalent to a permutation of the output signal that results from processing $\mathbf{x}$ with the filter instantiated on the shift operator $\mathcal{S}$. This is a consequence of the fact that the adjoint permutation $\mathbf{P}^T$ is the inverse of the permutation $\mathbf{P}$. Equivariance to permutations is important in graph signal processing and graphon signal processing. It implies processing that is independent of labeling.

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. 2) 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 $\sigma_\ell : \mathcal{M}_\ell \rightarrow \mathcal{M}_{\ell+1}$ that perform those operations of point-wise nonlinearity and pooling. Then, the output from the layer ℓ in the AlgNN is given by

$$\mathbf{x}_\ell = \sigma_\ell(\rho_\ell(a_\ell)\mathbf{x}_{\ell-1}) \quad (17)$$

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

$$\mathbf{x}_\ell = \Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_{\ell-1}, \mathcal{S}_{\ell-1}), \quad (18)$$

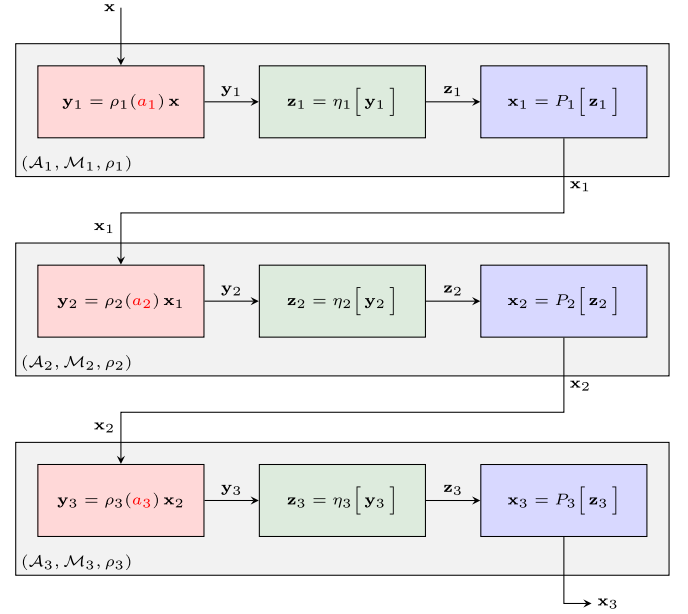


Fig. 2. Algebraic Neural Network $\Xi = \{(\mathcal{A}_\ell, \mathcal{M}_\ell, \rho_\ell)\}_{\ell=1}^3$ with three layers indicating how the input signal $\mathbf{x}$ is processed by Ξ and mapped into $\mathbf{x}_3$.

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 $\Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L)$ represents the total map associated to an AlgNN acting on a signal $\mathbf{x}$.

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

$$\mathbf{x}_\ell^f = \sigma_\ell \left(\sum_{g=1}^{F_\ell} \rho_\ell(a_\ell^{gf}) \mathbf{x}_{\ell-1}^g \right), \quad (19)$$

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.

Pooling: As stated in [24] 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.

We present some examples to clarify ideas.

Example 7 (CNNs in Discrete Time): 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 1. Consequently, the f th feature in layer ℓ is given by

$$\mathbf{x}_\ell^f = \sigma_\ell \left(\sum_{g=1}^{F_\ell} \sum_{k=1}^K h_{\ell k}^{gf} S_\ell^k \mathbf{x}_{\ell-1}^g \right), \quad (20)$$

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

Example 8 (Graph Neural Networks): In graph neural networks the algebraic signal model in each layer corresponds to the one discussed in example 2. Therefore, the f th feature in layer ℓ has the form

$$\mathbf{x}_\ell^f = \sigma_\ell \left(\sum_{g=1}^{F_\ell} \sum_{k=1}^K h_{\ell k}^{gf} \mathbf{S}_\ell^k \mathbf{x}_{\ell-1}^g \right), \quad (21)$$

where $\rho_\ell = S_\ell$. Here P_ℓ can be a dimensionality reduction operator or a zeroing operator that nullify components of the signal keeping its dimensionality. A common choice of the nonlinearity function is given by $\eta_\ell(u) = \max\{0, u\}$.

Example 9 (Group Neural Networks): In group neural networks the algebraic model is the same as specified in example 5. Therefore, the f th feature in layer ℓ is given by

$$\mathbf{x}_\ell^f = \sigma_\ell \left(\sum_{n=1}^{F_\ell} \sum_{u, h \in G_\ell} \mathbf{a}_{(\ell)}^{nf}(uh^{-1}) \mathbf{x}_{\ell-1}(h)u \right). \quad (22)$$

Where G_ℓ is the group associated to the ℓ th layer and $\mathbf{a}_{(\ell)}^{nf}$ are the coefficients of the filter associated to the feature f in layer ℓ . In this case $P_\ell : L^2(G_\ell) \rightarrow L^2(G_{\ell+1})$, 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 $\mathbb{R}^{|G_\ell|} \rightarrow \mathbb{R}^{|G_{\ell+1}|}$.

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 $a = p(g)$ we have that

$$\tilde{\rho}(a) = p_{\mathcal{M}}(\tilde{\rho}(g)) = p_{\mathcal{M}}(\tilde{\mathbf{S}}) = p(\tilde{\mathbf{S}}), \quad (23)$$

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

$$\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{T}(\mathbf{S}), \quad (24)$$

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 (24) and this perturbation propagates to the filter $\rho(a)$ according to (23). 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 justifiable by practical considerations. In the case of graph

signals a perturbation of the homomorphism models changes in the graph or errors in the measurement of edge weights. In the case of time signals, images, or groups, a perturbation of the homomorphism is an appropriate model of a diffeomorphism – a small warping of the domain. See Section V-D for more details.

Of the other components of an algebraic filter, the algebra and the vector space define the choice of operations and therefore are not naturally subject to perturbation. Perturbations of the input signal $\mathbf{x}$ are possible in practice but their theoretical analysis is simple. Filters are linear functions of the input and the nonlinear operations of AlgNNs are Lipschitz. Thus, algebraic filters and AlgNNs are readily shown to be Lipschitz stable to perturbations of the input $\mathbf{x}$.

A. Perturbation Models

In our subsequent analysis we consider perturbation models of the form

$$\mathbf{T}(\mathbf{S}) = \mathbf{T}_0 + \mathbf{T}_1 \mathbf{S}, \quad (25)$$

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 [7]. 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 (25) it is important to describe the commutativity of the shift operator $\mathbf{S}$ and the perturbation model operators $\mathbf{T}_r$. To that end, we write

$$\mathbf{S} \mathbf{T}_r = \mathbf{T}_{cr} \mathbf{S} + \mathbf{S} \mathbf{P}_r, \quad (26)$$

where $\mathbf{T}_{cr} = \sum_i \mu_i \mathbf{u}_i \langle \mathbf{u}_i, \cdot \rangle$, μ_i is the i th eigenvalue of $\mathbf{T}_r$, $\mathbf{u}_i$ is the i th eigenvector of $\mathbf{S}$, and $\langle \cdot, \cdot \rangle$ 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 δ according to

$$\|\mathbf{P}_r\|_F \leq \delta \|\mathbf{T}_r\|, \quad (27)$$

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 (27) can be bounded as we show in Proposition 8. The specifics of this bound are not central to the results of Section V. 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 $p : \text{End}(\mathcal{M})^m \rightarrow \text{End}(\mathcal{M})$, where $\text{End}(\mathcal{M})^m$ is the m -times cartesian product of $\text{End}(\mathcal{M})$. In this scenario we use the notation $p(\mathbf{S}) = p(\mathbf{S}_1, \dots, \mathbf{S}_m)$ and when considering the perturbation model in eqn. (25) acting on $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_m)$ we use the following notation $\mathbf{T}(\mathbf{S}) = (\mathbf{T}(\mathbf{S}_1), \dots, \mathbf{T}(\mathbf{S}_m))$ 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 $\mathcal{P} \subset \mathcal{A}$ which is denoted as $p \in \mathcal{P} \subset \mathcal{A}$, 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

$$\begin{aligned} & \|p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x}\| \leq \\ & \left[C_0 \sup_{\mathbf{S} \in \mathcal{S}} \|\mathbf{T}(\mathbf{S})\| + C_1 \sup_{\mathbf{S} \in \mathcal{S}} \|D_{\mathbf{T}}(\mathbf{S})\| + \mathcal{O}(\|\mathbf{T}(\mathbf{S})\|^2) \right] \|\mathbf{x}\|, \end{aligned} \quad (28)$$

for all $\mathbf{x} \in \mathcal{M}$. In (28) $D_{\mathbf{T}}(\mathbf{S})$ 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 (28) 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 (28). We know that this is true because unstable counterexamples are known in the case of graph signal processing [23] and the processing of time signals [1]. The best known example of an unstable filter is a high-pass filter in time when consider a dilation of the time line [6]. The same phenomenon is observed for graph signals when considering the dilation of graph shift operator [23].

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. To do so, we introduce the notion of frequency representation of the elements of an algebra as follows.

Definition 6 (Frequency Representation of a Filter): Consider an algebra $\mathcal{A}$ with a single generator g so that for all $a \in \mathcal{A}$ we can write $a = p(g)$. Let $\lambda \in \mathbb{F}$ be a variable taking values on the field $\mathbb{F}$. We say that $p(\lambda)$ is the frequency representation of the filter $a = p(g)$.

Notice that the frequency representation of the elements of the algebra $\mathcal{A}$ induces an isomorphism of algebras $\iota : \mathcal{A} \mapsto \mathcal{A}_{\mathbb{F}}$, where $\mathcal{A}_{\mathbb{F}}$ is obtained when the variables of elements in $\mathcal{A}$ are evaluated in $\mathbb{F}$. Then, we can characterize elements in $\mathcal{A}$ by means of the properties of their frequency representations. In what follows we introduce a definition used to characterize subsets of filters in algebras with a single generator that are relevant in our analysis.

Definition 7: Let $p : \mathbb{F} \rightarrow \mathbb{F}$ be the frequency representation of an element in an algebra with a single generator. Then, it is said that p is Lipschitz if there exists $L_0 > 0$ such that

$$|p(\lambda) - p(\mu)| \leq L_0 |\lambda - \mu| \quad (29)$$

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

$$\left| \lambda \frac{dp(\lambda)}{d\lambda} \right| \leq L_1 \text{ for all } \lambda. \quad (30)$$

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 .

We start our discussion on stability with a result for operators in algebraic models with a single generator. The result highlights 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 $\rho(g) = \mathbf{S} \in \text{End}(\mathcal{M})$. 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. (24). Then, for any $p \in \mathcal{A}$ we have

$$\|p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x}\| \leq \|\mathbf{x}\| (\|D_p(\mathbf{S})\{\mathbf{T}(\mathbf{S})\}\| + \mathcal{O}(\|\mathbf{T}(\mathbf{S})\|^2)) \quad (31)$$

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 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 act on the perturbation via the operator $D_p(\mathbf{S})$. Additionally, notice that eqn. (31) 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. (25). If $p \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$, then

$$\|D_p \mathbf{T}(\mathbf{S})\| \leq (1 + \delta) \left(L_0 \sup_{\mathbf{S}} \|\mathbf{T}(\mathbf{S})\| + L_1 \sup_{\mathbf{S}} \|D_{\mathbf{T}}(\mathbf{S})\| \right) \quad (32)$$

Proof: See Section VII-B. ■

It is worth pointing out that the constants involved in the upper bound of eqn. (32) 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. (25). Then,

if $p \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ 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 (32) from Theorem 2 into (31) from Theorem 1 and reorder terms. ■

B. Algebraic Filter Stability in Algebras with Multiple Generators

The stability results presented in previous subsection can be extended naturally to operators associated to representations of algebras with multiple generators. To do so, we introduce the notion of frequency representation of elements of algebras with multiple generators as follows.

Definition 8 (Frequency Representation of a Filter): Consider an algebra $\mathcal{A}$ with generators $g_1, \dots, g_m$ so that for all $a \in \mathcal{A}$ we can write $a = p(g_1, \dots, g_m)$. Let $\lambda_i \in \mathbb{F}$ be variables taking values on the field $\mathbb{F}$. We say that $p(\lambda_1, \dots, \lambda_m)$ is the frequency representation of the filter $a = p(g_1, \dots, g_m)$.

Similar to the scenario of algebras with a single generator, the frequency representation of the elements of $\mathcal{A}$ induces an isomorphism of algebras $\iota : \mathcal{A} \mapsto \mathcal{A}_{\mathbb{F}}$, where $\mathcal{A}_{\mathbb{F}}$ is obtained when the variables of elements in $\mathcal{A}$ are evaluated in $\mathbb{F}$. In this way we have a characterization of elements in $\mathcal{A}$ when considering the properties of their frequency representations.

We extend definitions introduced before to characterize frequency representations in multivariate algebras.

Definition 9: Let $p : \mathbb{F}^m \rightarrow \mathbb{F}$ be the frequency representation of an element in an algebra with m generators. Then, it is said that p is Lipschitz if there exists $L_0 > 0$ such that

$$|p(\lambda) - p(\mu)| \leq L_0 \|\lambda - \mu\| \quad (33)$$

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

$$\left| \lambda_i \frac{\partial p(\lambda)}{\partial \lambda_i} \right| \leq L_1 \quad \text{for all } i \in \{1, \dots, m\}, \quad (34)$$

where $\lambda = (\lambda_1, \dots, \lambda_m)$ and $\mu = (\mu_1, \dots, \mu_m)$.

With these notions 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 $\rho(g_i) = \mathbf{S}_i \in \text{End}(\mathcal{M})$ for all i . Let $\tilde{\rho}(g_i) = \tilde{\mathbf{S}}_i \in \text{End}(\mathcal{M})$ 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. (24). Then, for any $p \in \mathcal{A}$ we have

$$\begin{aligned} & \|p(\mathbf{S})\mathbf{x} - p(\tilde{\mathbf{S}})\mathbf{x}\| \leq \\ & \|\mathbf{x}\| \sum_{i=1}^m (\|D_{p|\mathbf{S}_i}(\mathbf{S})\mathbf{T}(\mathbf{S}_i)\| + \mathcal{O}(\|\mathbf{T}(\mathbf{S}_i)\|^2)) \end{aligned} \quad (35)$$

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. (35) 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. (25). Then, if $p \in \mathcal{A}_{L_0} \cap \mathcal{A}_{L_1}$ it holds that

$$\begin{aligned} & \|D_{p|\mathbf{S}_i}(\mathbf{S})\mathbf{T}(\mathbf{S}_i)\| \\ & \leq (1 + \delta) \left(L_0 \sup_{\mathbf{S}_i \in \mathcal{S}} \|\mathbf{T}(\mathbf{S}_i)\| + L_1 \sup_{\mathbf{S}_i \in \mathcal{S}} \|D_{\mathbf{T}}(\mathbf{S}_i)\| \right) \end{aligned} \quad (36)$$

Proof: See Section VII-B. ■

It is important to remark that the upper bound in eqn. (36) 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. (25). Then, if $p \in \mathcal{A}_{L_0} \cup \mathcal{A}_{L_1}$ 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. (36) from theorem 4 into eqn. (35) from theorem 3 and organizing 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. (25). 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 Ξ and $\tilde{\Xi}$ in the layer ℓ respectively, we have

$$\begin{aligned} & \left\| \Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \mathcal{S}_\ell) - \Phi(\mathbf{x}_{\ell-1}, \mathcal{P}_\ell, \tilde{\mathcal{S}}_\ell) \right\| \\ & \leq C_\ell (1 + \delta_\ell) \left(L_0^{(\ell)} \sup_{\mathcal{S}_\ell} \|\mathbf{T}^{(\ell)}(\mathcal{S}_\ell)\| \right. \\ & \quad \left. + L_1^{(\ell)} \sup_{\mathcal{S}_\ell} \|D_{\mathbf{T}^{(\ell)}}(\mathcal{S}_\ell)\| \right) \|\mathbf{x}_{\ell-1}\| \end{aligned} \quad (37)$$

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 Δ_ℓ 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. (25). Then, if $\Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L)$ and $\Phi(\mathbf{x}, \{\tilde{\mathcal{P}}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L)$ represent the mapping operators associated to Ξ and $\tilde{\Xi}$ respectively, we have

$$\begin{aligned} & \left\| \Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L) - \Phi(\mathbf{x}, \{\tilde{\mathcal{P}}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L) \right\| \\ & \leq \sum_{\ell=1}^L \Delta_\ell \left(\prod_{r=\ell}^L C_r \right) \left(\prod_{r=\ell+1}^L B_r \right) \left(\prod_{r=1}^{\ell-1} C_r B_r \right) \|\mathbf{x}\| \end{aligned} \quad (38)$$

where C_ℓ is the Lipschitz constant of σ_ℓ and B_ℓ is a bound on the filter's norm, $\|\rho_\ell(a)\| \leq B_\ell$. The functions Δ_ℓ are given by

$$\Delta_\ell = (1 + \delta_\ell) \left(L_0^{(\ell)} \sup_{\mathbf{S}_\ell} \|\mathbf{T}^{(\ell)}(\mathbf{S}_\ell)\| + L_1^{(\ell)} \sup_{\mathbf{S}_\ell} \|D_{\mathbf{T}^{(\ell)}}(\mathbf{S}_\ell)\| \right) \quad (39)$$

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 [7] 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.

The bound in (38) exhibits an exponential dependency on the Lipschitz constants C_ℓ and the maximum filter norms B_ℓ . This dependency can be avoided if we normalize the nonlinearities so that $C_\ell = 1$ and the filters so that $B_\ell = 1$. Their presence in (38) highlights that if the filters and nonlinearities amplify signals, they may amplify errors as well.

Remark 3: It is important to highlight the fact that the perturbation model in eqn. (25) 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 signal models 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.

Remark 4: It is worth noticing that the role of the Fréchet derivative $\mathbf{D}_{p|\mathbf{S}}(\mathbf{S})$ raises naturally when the the norm of the difference between an operator $p(\mathbf{S})$ and its perturbed version $p(\tilde{\mathbf{S}})$ is considered (see Section VII-A), and this is a direct consequence of the definition of such derivative and the type of perturbation considered (see eqn. (49) and eqn. (50)). Then, as long as the perturbed shifts considered can be modeled as $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{T}(\mathbf{S})$, i.e. the perturbation is added to the unperturbed shift, the operator that acts on $\mathbf{T}(\mathbf{S})$ is always the Fréchet derivative of the filter.

D. Implications for particular signal models

In this subsection we show the implication of the stability results for particular signal models.

Graph Neural Networks (GNNs) 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 absolute perturbation model considered in [7] is the perturbation model where $\mathbf{T}(\mathbf{S}) = \mathbf{T}_0$. Therefore the stability bound for graph filters translates into

$$\|p(\mathbf{S}) - p(\tilde{\mathbf{S}})\| \leq (1 + \delta)L_0\|\mathbf{T}_0\| + \mathcal{O}(\|\mathbf{T}_0\|^2), \quad (40)$$

which is a scaled version of the result in [7] (Theorem 1). Additionally, $\delta \leq \hat{\delta}\sqrt{N}$ where $\hat{\delta}$ is the non commutativity constant used in [7] which depends on the difference between the eigenvectors of $\mathbf{S}$ and $\mathbf{T}_r$ – please see Appendix A where the formal connection between δ and $\hat{\delta}$ is stated. Notice that $\sup_{\mathbf{S}} \|\mathbf{T}(\mathbf{S})\| = \|\mathbf{T}_0\|$.

A relative model can be obtained considering $\mathbf{T}(\mathbf{S}) = \mathbf{T}_1\mathbf{S}$, and in that case the stability bounds are given according to

$$\|p(\mathbf{S}) - p(\tilde{\mathbf{S}})\| \leq (1 + \delta)L_1\|\mathbf{T}_1\| + \mathcal{O}(\|\mathbf{T}(\mathbf{S})\|^2), \quad (41)$$

which is a scaled version of the bound obtained in [7]. Notice that $\sup_{\mathbf{S}} \|\mathbf{D}_{\mathbf{T}}(\mathbf{S})\| = \|\mathbf{T}_1\|$. Like in the previous scenario $\delta \leq \sqrt{N}\hat{\delta}$ where $\hat{\delta}$ is the non commutativity constant used in [7] – please see Appendix A. It is also important to remark that the stability of bounds derived in [25] for graph scattering transforms are rooted in the fact that wavelet graph filters are stable, and as a consequence the stability bounds are scaled versions of the ones derived for graph filters.

CNNs with DTSP In discrete time signal processing (DTSP) the shift operator is the discrete time shift S . The processing induced by (5) 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 (28) 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 $\tilde{S}$. The stability bound in this scenario is given by

$$\begin{aligned} \|p(S)\mathbf{x} - p(\tilde{S})\mathbf{x}\| &\leq \left[L_0(1 + \delta) \sup_{S \in \mathcal{S}} \|\mathbf{T}(S)\| \right. \\ &\quad \left. + L_1(1 + \delta) \sup_{S \in \mathcal{S}} \|D_{\mathbf{T}}(S)\| + \mathcal{O}(\|\mathbf{T}(S)\|^2) \right] \|\mathbf{x}\|. \end{aligned} \quad (42)$$

Notice that although results in eqn. (42) are different from those in [1], they exhibit similarities. This is expected since the right hand side of eqn. (42) measures the size of $\mathbf{T}(S)$, which is a diffeomorphism acting on the space of admissible shift operators. The bounds derived in [1] consider diffeomorphisms acting on $\mathbb{R}^n$ which is the domain of the signals. Additionally, notice that since the operators considered in [1] are shift invariant the term associated to the absolute norm of the deformation is not present in the bounds. It is also worth pointing out that the convolutions we consider in the DTSP model are attributed to a polynomial algebra. While the convolutions considered in [1] are defined considering functions in $L_2(\mathbb{R}^n)$ and filters in $L_1(\mathbb{R})$, a scenario that requires the use of a non polynomial algebra.

Graphon Neural Networks Similar to the case of GNNs the graphon $W(u, v)$ is a limit object that represents a family of random graphs. The perturbed graphon $\tilde{W}(u, v)$ represents a different family of random graphs. The perturbation of the graphon generates a corresponding perturbation of the shift operator defined in (13). If the condition stated in (28) is satisfied, then filtering graphon signals using the perturbed shift operator associated to $\tilde{W}(u, v)$ 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 $W(u, v)$. For instance, if the perturbation considered is additive we have

$$\|p(S) - p(\tilde{S})\| \leq (1 + C)L_0\|\mathbf{T}_0\| + \mathcal{O}(\|\mathbf{T}_0\|^2), \quad (43)$$

where S is the graphon shift operator indicated in eqn. (6) and C is a constant associated to the eigenvalue and eigenvector spreading of the graphon operator.

Group Neural Network Similar to the case of DSP, the filters in (11) 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 (28) 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. Notice that when considering the representations of finite commutative groups the analysis of stability is the same as in the case of an architecture based on a DSP model, therefore the stability bounds to perturbations are given by eqn. (42).

VI. SPECTRAL OPERATORS

Part of our proofs on the stability of AlgNN rely on the notion of spectral or Fourier decompositions associated to the realization of algebraic filters. In this section we discuss the notion of spectrum for general operators associated to algebraic signal models. Such notions of spectral decompositions are a natural generalization of the well established notions of spectrum used in GNNs and CNNs. To do so we elaborate about the concepts of irreducible and indecomposable subrepresentations, which generalize the notions of decompositions in terms of eigenvectors and eigenvalues [8], [26], [27].

We will 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 eigenvalues¹ 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 start introducing the notion of subrepresentation.

Definition 10: 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 $(\mathcal{M}, \rho)$.

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 11: 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** $\phi : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is a linear operator which commutes with the action of $\mathcal{A}$, i.e.

$$\phi(\rho_1(a)v) = \rho_2(a)\phi(v). \quad (44)$$

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

Notice from definition 11 a substantial difference between the concepts of isomorphism of vector spaces and isomorphism of representations. In the first case we can consider that two arbitrary vector spaces of the same dimension (finite) are isomorphic, while for representations that condition is required but still the condition in eqn. (44) must be satisfied. For instance, as pointed out in [28] 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

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 $\mathcal{A}$.

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 12: 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) [29]. The following theorem provides the basic building block for the decomposition of finite dimensional representations.

Theorem 7 (Krull-Schmit, [30]): 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 $(\bigoplus_{r=1}^r V_i, \rho) \cong (\bigoplus_{j=1}^s W_j, \gamma)$ 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 $(V_i, \rho_i) \cong (W_{\pi(j)}, \gamma_{\pi(j)})$ [30]. 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 [28], [30], 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 $\text{End}(\mathcal{M})$ [30]. 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 13 ([30]): 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 $m(\mathcal{U}, V)$ [30]. 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 $(V, \rho_V) = \text{soc}\{V\}$ [30]. The connection between $\text{soc}\{V\}$ and $V(\mathcal{U})$ is given by the following proposition.

Proposition 1 (Proposition 1.31 [30]): Let $(V, \rho_V) \in \text{Rep}\{\mathcal{A}\}$. Then $\text{soc}\{V\} = \bigoplus_{S \in \text{Irr}\{\mathcal{A}\}} V(S)$.

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

$$\text{soc}\{V\} \cong \bigoplus_{S \in \text{Irr}\{\mathcal{A}\}} S^{\oplus m(\mathcal{U}, V)}. \quad (45)$$

Equation (45) provides the building block for the definition of Fourier decompositions in algebraic signal processing [11]. With all these concepts at hand we are ready to introduce the following definition.

Definition 14 (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

$$(\mathcal{M}, \rho) \cong \bigoplus_{(\mathcal{U}_i, \phi_i) \in \text{Irr}\{\mathcal{A}\}} (\mathcal{U}_i, \phi_i)^{\oplus m(\mathcal{U}_i, \mathcal{M})} \quad (46)$$

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

$$\Delta : \mathcal{M} \rightarrow \bigoplus_{(\mathcal{U}_i, \phi_i) \in \text{Irr}\{\mathcal{A}\}} (\mathcal{U}_i, \phi_i)^{\oplus m(\mathcal{U}_i, \mathcal{M})} \quad \mathbf{x} \mapsto \hat{\mathbf{x}} \quad (47)$$

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. (46) 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 5: The spectral representation of an operator indicated as $\phi_i(a)\hat{\mathbf{x}}(i)$ and eqns. (46) and (47) highlight one important fact that is essential for the discussion of the results in Section VII. For a completely reducible representation $(\mathcal{M}, \rho) \in \text{Rep}\{\mathcal{A}\}$ 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.(46). 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 [7] 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 10 (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

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

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 $\phi_i(t) = e^{-\frac{2\pi j i}{N} t}$ is the eigenvalue associated to $\mathbf{u}_i$. Here $\hat{\mathbf{x}}$ represents the DFT of $\mathbf{x}$.

Example 11 (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

$$\begin{aligned}\rho(a)\mathbf{x} &= \sum_{i=1}^N \phi_i \left(\sum_{k=0}^{K-1} h_k t^k \right) \hat{\mathbf{x}}(i) \mathbf{u}_i \\ &= \sum_{i=1}^N \sum_{k=0}^{K-1} h_k \phi_i(t)^k \hat{\mathbf{x}}(i) \mathbf{u}_i = \sum_{i=1}^N \sum_{k=0}^{K-1} h_k \lambda_i^k \hat{\mathbf{x}}(i) \mathbf{u}_i\end{aligned}\quad (48)$$

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 $\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x}$ [23].

Example 12 (Group signal processing): Considering the Fourier decomposition on general groups [16]–[18], we obtain the spectral representation of the algebraic filters as

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

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

Example 13 (Graphon signal processing): According to the spectral theorem [31], [32], it is possible to represent the action of a compact normal operator S as $S\mathbf{x} = \sum_i \lambda_i \langle \varphi_i, \mathbf{x} \rangle \varphi_i$ where λ_i and φ_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

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

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

VII. PROOF OF THEOREMS

Let us start defining some notation. Let $\pi_{a_1, \dots, a_r}(\mathbf{A}_1, \dots, \mathbf{A}_r)$ be the operator that represents

the sum of all the products of the operators $\mathbf{A}_1, \dots, \mathbf{A}_r$ that appear $a_1, a_2, \dots, a_r$ times respectively. For instance, $\pi_{2,1}(\mathbf{A}, \mathbf{B}) = \mathbf{AAB} + \mathbf{ABA} + \mathbf{BAA}$. 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 $b < a$. In what follows $\|\cdot\|$ represents the ℓ_2 norm and $\|\cdot\|_F$ the Frobenius norm.

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 $D_p : \text{End}(\mathcal{M})^m \rightarrow \text{End}(\mathcal{M})$ such that [33], [34]

$$\lim_{\|\xi\| \rightarrow 0} \frac{\|p(\mathbf{S} + \xi) - p(\mathbf{S}) - D_p(\mathbf{S})\{\xi\}\|}{\|\xi\|} = 0 \quad (49)$$

which in Landau notation can be written as

$$p(\mathbf{S} + \xi) = p(\mathbf{S}) + D_p(\mathbf{S})\{\xi\} + o(\|\xi\|). \quad (50)$$

Calculating the norm in eqn. (50) and applying the triangle inequality we have:

$$\|p(\mathbf{S} + \xi) - p(\mathbf{S})\| \leq \|D_p(\mathbf{S})\{\xi\}\| + \mathcal{O}(\|\xi\|^2) \quad (51)$$

for all $\xi = (\xi_1, \dots, \xi_m) \in \text{End}(\mathcal{M})^m$. Now, taking into account the properties of a Fréchet derivative for a function of multiple variables (see [35] pages 69-70) we have

$$\|D_p(\mathbf{S})\{\xi\}\| \leq \sum_{i=1}^m \|D_{p|S_i}(\mathbf{S})\{\xi_i\}\| \quad (52)$$

and therefore

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

where $D_{p|S_i}(\mathbf{S})$ is the partial Frechet derivative of $p(\mathbf{S})$ on S_i . Then, taking into account that

$$\|p(\mathbf{S} + \xi)\mathbf{x} - p(\mathbf{S})\mathbf{x}\| \leq \|\mathbf{x}\| \|p(\mathbf{S} + \xi) - p(\mathbf{S})\| \quad (53)$$

and selecting $\xi_i = \mathbf{T}(S_i)$ we complete the proof. ■

B. Proof of Theorem 2 and Theorem 4

Proof: Taking into account the definition of the Fréchet derivative of p on S_i (see Appendix B) we have

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

and re-organizing terms we have

$$\|D_{p|S_i}(\mathbf{S})\{\mathbf{T}(S_i)\}\| = \left\| \sum_{\ell=1}^{\infty} S_i^{\ell-1} \mathbf{T}(S_i) \sum_{k_i=\ell}^{\infty} \mathbf{A}_{k_i} S_i^{k_i-\ell} \right\|. \quad (54)$$

Taking into account eqn. (26), it follows that

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

$$+ \sum_{\ell=1}^{\infty} (\mathbf{T}_{1c,i} \mathbf{S}_i^{\ell} + \mathbf{S}_i^{\ell-1} \mathbf{P}_{1,i} \mathbf{S}_i) \sum_{k=\ell}^{\infty} \mathbf{A}_{k_i} \mathbf{S}_i^{k_i-\ell} \Big\| . \quad (55)$$

Applying the triangle inequality and distributing the sum we have

$$\begin{aligned} \|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{T}(\mathbf{S}_i)\}\| &\leq \left\| \mathbf{T}_{0c,i} \sum_{\ell=1}^{\infty} \sum_{k_i=\ell}^{\infty} \mathbf{S}_i^{k_i-1} \mathbf{A}_{k_i} \right\| \\ &+ \|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{P}_{0,i}\}\| + \left\| \mathbf{T}_{1c,i} \sum_{\ell=1}^{\infty} \sum_{k_i=\ell}^{\infty} \mathbf{S}_i^{k_i} \mathbf{A}_{k_i} \right\| \\ &+ \|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{P}_{1,i} \mathbf{S}_i\}\| \end{aligned} \quad (56)$$

Now, we analyze term by term in eqn. (56). For the first term we take into account that

$$\sum_{\ell=1}^{\infty} \sum_{k_i=\ell}^{\infty} \mathbf{S}_i^{k_i-1} \mathbf{A}_{k_i} = \sum_{k_i=1}^{\infty} k_i \mathbf{A}_{k_i} \mathbf{S}_i^{k_i-1} \quad (57)$$

and we apply the product norm property taking into account that the filters belong to $\mathcal{A}_{L_0}$, which leads to

$$\begin{aligned} &\left\| \mathbf{T}_{0c,i} \sum_{\ell=1}^{\infty} \sum_{k_i=\ell}^{\infty} \mathbf{S}_i^{k_i-1} \mathbf{A}_{k_i} \right\| \\ &\leq \|\mathbf{T}_{0c,i}\| \left\| \sum_{k_i=1}^{\infty} k_i \mathbf{A}_{k_i} \mathbf{S}_i^{k_i-1} \right\| \leq L_0 \|\mathbf{T}_{0,i}\|. \end{aligned} \quad (58)$$

For the second term in eqn. (56) we take into account that the Fréchet derivative acting on $\mathbf{P}_{0,i}$ can be equivalently expressed as a linear operator acting on the left of a vectorized version of $\mathbf{P}_{0,i}$ (see [36] pages 61 and 331). Then,

$$\|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{P}_{0,i}\}\| \leq L_0 \|\mathbf{P}_{0,i}\|_F \quad (59)$$

and with the fact that $\|\mathbf{P}_{0,i}\|_F \leq \delta \|\mathbf{T}_{0,i}\|$, we have

$$\|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{P}_{0,i}\}\| \leq L_0 \delta \|\mathbf{T}_{0,i}\|. \quad (60)$$

For the third term in eqn. (56), we take into account that

$$\sum_{\ell=1}^{\infty} \sum_{k_i=\ell}^{\infty} \mathbf{S}_i^{k_i} \mathbf{A}_{k_i} = \sum_{k_i=1}^{\infty} k_i \mathbf{A}_{k_i} \mathbf{S}_i^{k_i}, \quad (61)$$

and we apply the norm product property taking into account that the filters belong to $\mathcal{A}_{L_1}$, which leads to

$$\begin{aligned} &\left\| \mathbf{T}_{1c,i} \sum_{\ell=1}^{\infty} \sum_{k_i=\ell}^{\infty} \mathbf{S}_i^{k_i} \mathbf{A}_{k_i} \right\| \\ &\leq \|\mathbf{T}_{1c,i}\| \left\| \sum_{k_i=1}^{\infty} k_i \mathbf{A}_{k_i} \mathbf{S}_i^{k_i} \right\| \leq L_1 \|\mathbf{T}_{1,i}\|. \end{aligned} \quad (62)$$

Finally, for the fourth term we use the notation $\tilde{D}(\mathbf{S})\{\mathbf{P}_{1,i}\} = D_{p|\mathbf{S}_i}(\mathbf{S})\{\mathbf{P}_{1,i} \mathbf{S}_i\}$. We start pointing out that (see [36] pages 61 and 331) the eigenvalues of the operator $\tilde{D}(\mathbf{S})$ represented as ζ_{pq} are given by

$$\zeta_{pq} = \begin{cases} \frac{p(\lambda_p) - p(\lambda_q)}{\lambda_p - \lambda_q} \lambda_q & \text{if } \lambda_p \neq \lambda_q \\ \lambda_p p'(\lambda_p) & \text{if } \lambda_p = \lambda_q \end{cases}. \quad (63)$$

Then, taking into account that the filters belong to $\mathcal{A}_{L_1}$ we have $\|\tilde{D}(\mathbf{S})\| \leq L_1$, and therefore

$$\|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{P}_{1,i} \mathbf{S}_i\}\| = \|\tilde{D}(\mathbf{S}) \{\mathbf{P}_{1,i}\}\| \leq L_1 \|\mathbf{P}_{1,i}\|_F \quad (64)$$

Additionally, with $\|\mathbf{P}_{1,i}\|_F \leq \delta \|\mathbf{T}_{1,i}\|$ it follows that

$$\|D_{p|\mathbf{S}_i}(\mathbf{S}) \{\mathbf{P}_{1,i} \mathbf{S}_i\}\| \leq L_1 \delta \|\mathbf{T}_{1,i}\| \quad (65)$$

Putting all these results together into eqn. (56) we reach

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

C. Proof of Theorems 5 and 6

1) *Proof of Theorem 5: Proof:* Taking into account eqns. (17), and (18) and the fact that the maps σ_{ℓ} are Lipschitz with constant C_{ℓ} we have that

$$\|\sigma_{\ell}(p(\mathbf{S}_{\ell}) \mathbf{x}_{\ell-1}) - \sigma_{\ell}(p(\tilde{\mathbf{S}}_{\ell}) \mathbf{x}_{\ell-1})\| \leq C_{\ell} \Delta_{\ell} \|\mathbf{x}_{\ell-1}\|, \quad (66)$$

where $\Delta_{\ell} = \|p(\mathbf{S}_{\ell}) - p(\tilde{\mathbf{S}}_{\ell})\|$, 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 $\varphi_f(\ell, g) = \rho_{\ell}(\xi^{fg})$ denote the image of the filter $\xi^{fg} \in \mathcal{A}_{\ell}$ that process the g th feature coming from the layer $\ell - 1$ and that is associated to f th 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 g th feature in the layer $\ell - 1$. Then, we have that:

$$\begin{aligned} \|\mathbf{x}_{\ell}^f - \tilde{\mathbf{x}}_{\ell}^f\| &\leq \left\| \sigma_{\ell-1} \sum_{g_{\ell-1}} \varphi_{g_{\ell}}(\ell - 1, g_{\ell-1}) \sigma_{\ell-2} \right. \\ &\quad \sum_{g_{\ell-2}} \varphi_{g_{\ell-1}}(\ell - 2, g_{\ell-2}) \cdots \sigma_1 \sum_{g_1} \varphi_{g_2} \mathbf{x} - \\ &\quad \sigma_{\ell-1} \sum_{g_{\ell-1}} \tilde{\varphi}_{g_{\ell}}(\ell - 1, g_{\ell-1}) \sigma_{\ell-2} \sum_{g_{\ell-2}} \tilde{\varphi}_{g_{\ell-1}}(\ell - 2, g_{\ell-2}) \\ &\quad \left. \cdots \sigma_1 \sum_{g_1} \tilde{\varphi}_{g_2}(1, g_1) \mathbf{x} \right\|. \end{aligned} \quad (67)$$

In order to expand eqn. (67) we start pointing out that:

$$\begin{aligned} A_{\ell+1} \sigma_{\ell}(a) - \tilde{A}_{\ell+1} \sigma_{\ell}(\tilde{a}) &= \\ (A_{\ell+1} - \tilde{A}_{\ell+1}) \sigma_{\ell}(a) + \tilde{A}_{\ell+1} (\sigma_{\ell}(a) - \sigma_{\ell}(\tilde{a})) \end{aligned} \quad (68)$$

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

$$\sum_{g_k} \|\alpha - \tilde{\alpha}\| \leq \sum_{g_k} \left(\Delta_k \|\sigma_{k-1}(\alpha)\| + B_k C_{k-1} \|\beta - \tilde{\beta}\| \right) \quad (69)$$

$$\sum_{g_k} \|\beta - \tilde{\beta}\| \leq \sum_{g_k} \sum_{g_{k-1}} \|\alpha - \tilde{\alpha}\| \quad (70)$$

$$\sum_{g_k} \|\sigma_{k-1}(\alpha)\| \leq \left(\prod_{r=1}^k F_r \right) \left(\prod_{r=1}^{k-1} C_r B_r \right) \|\mathbf{x}\| \quad (71)$$

where α and $\tilde{\alpha}$ represent sequences of symbols in eqn. (67) 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 Δ_ℓ 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. (69), (70) and (71) we have:

$$\begin{aligned} \|\mathbf{x}_L^f - \tilde{\mathbf{x}}_L^f\| &\leq \sum_{\ell=1}^L \Delta_\ell \left(\prod_{r=\ell}^L C_r \right) \left(\prod_{r=\ell+1}^L B_r \right) \\ &\left(\prod_{r=\ell}^{L-1} F_r \right) \left(\prod_{r=1}^{\ell-1} C_r F_r B_r \right) \|\mathbf{x}\|, \end{aligned} \quad (72)$$

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

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

we have

$$\begin{aligned} &\left\| \Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\mathcal{S}_\ell\}_1^L) - \Phi(\mathbf{x}, \{\mathcal{P}_\ell\}_1^L, \{\tilde{\mathcal{S}}_\ell\}_1^L) \right\| \\ &\leq \sqrt{F_L} \sum_{\ell=1}^L \Delta_\ell \left(\prod_{r=\ell}^L C_r \right) \left(\prod_{r=\ell+1}^L B_r \right) \left(\prod_{r=\ell}^{L-1} F_r \right) \\ &\left(\prod_{r=1}^{\ell-1} C_r F_r B_r \right) \|\mathbf{x}\| \end{aligned} \quad (73)$$

VIII. DISCUSSION

The mathematical form of the notion of stability introduced in definition (5), eqn. (28) 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 $\text{End}(\mathcal{M})$. 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 norm of the Fréchet derivative of the filters $\|D_{p|\mathbf{S}}(\mathbf{S})\|$. Those filters with slow variation and low selectivity

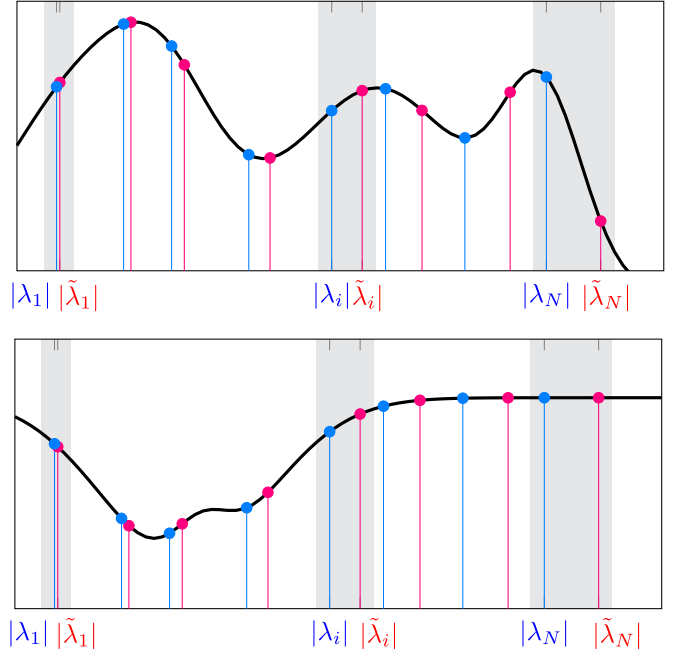


Fig. 3. Filter properties and stability for algebraic operators considering algebras with a single generator. (Top) We depict a Lipschitz filter where it is possible to see that an arbitrary degree of selectivity can be achieved in any part of the spectrum. (bottom) We depict a Lipschitz integral filter 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.

will be associated with a low value of $\|D_{p|\mathbf{S}}(\mathbf{S})\|$ while a filter that high variation will lead to large values of $\|D_{p|\mathbf{S}}(\mathbf{S})\|$. This is also reflected in 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 $\mathcal{A}_{L_0}$ and $\mathcal{A}_{L_1}$. The smaller the value of L_0, L_1 the more stable the operators but the less selectivity we have. In Fig. (3) the properties in frequency of Lipschitz 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 relocater 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 Lipschitz functions, but substantial differences may exist if

$\dim(\text{End}(\mathcal{M})) = \infty$ even if the functions are Lipschitz [33], [34].

IX. CONCLUSION

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 homomorphisms in $\text{End}(\mathcal{M})$ 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 used in this work.

An interesting and relevant future research direction is to analyze stability of operators in signal models with non commutative algebras. This is important since we do not have shift invariance, and consequently there is the question of how this affects the stability properties and the constants in a stability bound if it exists. Another essential question to solve is how the ASP theory can be extended to consider stability of convolutional operators in signal models where the algebra is not of polynomial type. This has implications when considering convolutions with functions in $L_2(\mathbb{R}^n)$, where the algebra is $L_1(\mathbb{R}^n)$ and the notion of generator set proposed in [8] is insufficient/inadequate to capture the whole structure of the algebra.

APPENDIX A PERTURBATION MODEL

Theorem 8: Let $\mathbf{T}_r, \mathbf{T}_{cr}$ as specified in eqn. (26) for the perturbation model with $\|\mathbf{P}_r\|_F \leq \delta \|\mathbf{T}_r\|$. Let $\mathbf{e}_i$ a orthonormal basis, $(\lambda_i, \mathbf{v}_i)$ the eigenpairs of $\mathbf{S}$ and $(\mu_i, \mathbf{u}_i)$ the eigenpairs of $\mathbf{T}_r$. Then

$$\delta \leq \sqrt{N} \hat{\delta} \quad (74)$$

where

$$\hat{\delta} = (\|\mathbf{T}_u - \mathbf{T}_v\| + 1)^2 - 1 \quad (75)$$

and

$$\mathbf{T}_v = \left(\sum_i \mathbf{e}_i \langle \mathbf{v}_i, \cdot \rangle \right), \quad \mathbf{T}_u = \left(\sum_i \mathbf{e}_i \langle \mathbf{u}_i, \cdot \rangle \right). \quad (76)$$

The terms $\langle \mathbf{u}_i, \cdot \rangle$ and $\langle \mathbf{v}_i, \cdot \rangle$ indicate the inner product operators with $\mathbf{u}_i$ and $\mathbf{v}_i$ respectively.

Proof: For our analysis we consider the following operators

$$\mathbf{T}_v^* = \left(\sum_i \mathbf{v}_i \langle \cdot, \mathbf{e}_i \rangle \right), \quad \mathbf{T}_u^* = \left(\sum_i \mathbf{u}_i \langle \cdot, \mathbf{e}_i \rangle \right), \quad (77)$$

$$\mathbf{T}_\mu = \left(\sum_i \mu_i \mathbf{e}_i \langle \mathbf{e}_i, \cdot \rangle \right), \quad (78)$$

and we remark that

$$\mathbf{T}_r = \mathbf{T}_u^* \mathbf{T}_\mu \mathbf{T}_u, \quad \mathbf{T}_{cr} = \mathbf{T}_v^* \mathbf{T}_\mu \mathbf{T}_v \quad (79)$$

with

$$\|\mathbf{T}_u^*\| = \|\mathbf{T}_u\| = \|\mathbf{T}_v^*\| = \|\mathbf{T}_v\| = 1 \quad (80)$$

and

$$\|\mathbf{T}_\mu\| = \|\mathbf{T}_r\|. \quad (81)$$

Now, we start taking into account that $\mathbf{T}_r$ can be rewritten as

$$\begin{aligned} \mathbf{T}_r &= \mathbf{T}_v^* \mathbf{T}_\mu \mathbf{T}_v + (\mathbf{T}_u^* - \mathbf{T}_v^*) \mathbf{T}_\mu (\mathbf{T}_u - \mathbf{T}_v) \\ &\quad + \mathbf{T}_v^* \mathbf{T}_\mu (\mathbf{T}_u - \mathbf{T}_v) + (\mathbf{T}_u^* - \mathbf{T}_v^*) \mathbf{T}_\mu \mathbf{T}_v. \end{aligned} \quad (82)$$

Then, taking into account that $\mathbf{T}_r = \mathbf{T}_{cr} + \mathbf{P}_r$ we have that

$$\begin{aligned} \mathbf{P}_r &= (\mathbf{T}_u^* - \mathbf{T}_v^*) \mathbf{T}_\mu (\mathbf{T}_u - \mathbf{T}_v) \\ &\quad + \mathbf{T}_v^* \mathbf{T}_\mu (\mathbf{T}_u - \mathbf{T}_v) + (\mathbf{T}_u^* - \mathbf{T}_v^*) \mathbf{T}_\mu \mathbf{T}_v. \end{aligned} \quad (83)$$

Computing the norm on both sides of eqn. (83) and applying the triangular inequality and the operator norm property it follows that

$$\begin{aligned} \|\mathbf{P}_r\| &\leq \|\mathbf{T}_u^* - \mathbf{T}_v^*\| \|\mathbf{T}_\mu\| \|\mathbf{T}_u - \mathbf{T}_v\| \\ &\quad + \|\mathbf{T}_v^*\| \|\mathbf{T}_\mu\| \|\mathbf{T}_u - \mathbf{T}_v\| + \|\mathbf{T}_u^* - \mathbf{T}_v^*\| \|\mathbf{T}_\mu\| \|\mathbf{T}_v\|. \end{aligned} \quad (84)$$

Taking into account the expressions in eqn. (80), eqn. (81), and the fact that $\|\mathbf{T}_u^* - \mathbf{T}_v^*\| = \|\mathbf{T}_u - \mathbf{T}_v\|$, the eqn. (84) turns into

$$\|\mathbf{P}_r\| \leq \|\mathbf{T}_u - \mathbf{T}_v\|^2 \|\mathbf{T}_r\| + 2 \|\mathbf{T}_r\| \|\mathbf{T}_u - \mathbf{T}_v\| \quad (85)$$

which finally can be written as

$$\|\mathbf{P}_r\| \leq \|\mathbf{T}_r\| \left((\|\mathbf{T}_u - \mathbf{T}_v\| + 1)^2 - 1 \right). \quad (86)$$

Now, from the relationship between the Frobenius norm and the ℓ_2 -norm we know that

$$\frac{\|\mathbf{P}_r\|_F}{\sqrt{N}} \leq \|\mathbf{P}_r\| \quad (87)$$

Combining eqn. (87) and eqn. (86) it follows that

$$\|\mathbf{P}_r\|_F \leq \sqrt{N} \left((\|\mathbf{T}_u - \mathbf{T}_v\| + 1)^2 - 1 \right) \|\mathbf{T}_r\|. \quad (88)$$

Notice that the term $\|\mathbf{T}_u - \mathbf{T}_v\|$ is a measure of the difference between the eigenvectors of $\mathbf{S}$ and the eigenvectors of $\mathbf{T}_r$.

APPENDIX B FRECHET DERIVATIVE $D_{p|S_i}(\mathbf{S})$

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

Then, it follows that

$$p(\mathbf{S} + \boldsymbol{\xi}) - p(\mathbf{S}) = \sum_{k_i=0}^{\infty} (\mathbf{S}_i + \boldsymbol{\xi}_i)^{k_i} \mathbf{A}_{k_i} - \sum_{k_i=0}^{\infty} \mathbf{S}_i^{k_i} \mathbf{A}_{k_i} \quad (89)$$

for $\boldsymbol{\xi} = (\mathbf{0}, \dots, \boldsymbol{\xi}_i, \dots, \mathbf{0})$. Considering the expansion $(\mathbf{S}_i + \boldsymbol{\xi}_i)^{k_i} = \mathbf{S}_i^{k_i} + \boldsymbol{\xi}_i^{k_i} + \sum_{r=1}^{k_i-1} \pi_{r, k_i-r}(\mathbf{S}_i, \boldsymbol{\xi}_i)$ for $k_i \geq 2$, eqn. (89) takes the form

$$p(\mathbf{S} + \boldsymbol{\xi}) - p(\mathbf{S}) = \sum_{k_i=1}^{\infty} \sum_{r=1}^{k_i-1} \pi_{r, k_i-r}(\boldsymbol{\xi}_i, \mathbf{S}_i) \mathbf{A}_{k_i} + \sum_{k_i=1}^{\infty} \boldsymbol{\xi}_i^{k_i} \mathbf{A}_{k_i}. \quad (90)$$

Separating the linear terms on $\boldsymbol{\xi}_i$ eqn. (90) leads to

$$p(\mathbf{S} + \boldsymbol{\xi}) - p(\mathbf{S}) = \sum_{k_i=1}^{\infty} \pi_{1, k_i-1}(\boldsymbol{\xi}_i, \mathbf{S}_i) \mathbf{A}_{k_i} + \sum_{k_i=2}^{\infty} \sum_{r=2}^{k_i-1} \pi_{r, k_i-r}(\boldsymbol{\xi}_i, \mathbf{S}_i) \mathbf{A}_{k_i} + \sum_{k_i=2}^{\infty} \boldsymbol{\xi}_i^{k_i} \mathbf{A}_{k_i}. \quad (91)$$

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

$$D_{p|S_i}(\mathbf{S}) \{\boldsymbol{\xi}_i\} = \sum_{k_i=1}^{\infty} \pi_{1, k_i-1}(\boldsymbol{\xi}_i, \mathbf{S}_i) \mathbf{A}_{k_i} \quad (92)$$

REFERENCES

- [1] S. Mallat, "Group invariant scattering," *Commun. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, 2012.
- [2] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.
- [3] J. B. Estrach, A. Szlam, and Y. LeCun, "Learning stable group invariant representations with convolutional networks," in *Proc. 1st Int. Conf. Learn. Representations*, 2013, pp. 1–4.
- [4] A. Bietti and J. Mairal, "Invariance and stability of deep convolutional representations," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc, 2017, pp. 6211–6221.
- [5] D. Zou and G. Lerman, "Graph convolutional neural networks via scattering," *Appl. Comput. Harmon. Anal.*, vol. 49, no. 3, pp. 1046–1074, 2020.
- [6] F. Gama, J. B. Estrach, and A. Ribeiro, "Diffusion scattering transforms on graphs," in *Proc. 7th Int. Conf. Learn. Representations*, pp. 1–17, 2019.
- [7] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 5680–5695, 2020.
- [8] M. Püschel and M. F. José Moura, "Algebraic signal processing theory," 2006, *arXiv:abs/cs/0612077*.
- [9] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, Aug. 2017.
- [10] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Pearson, 2014.
- [11] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-D time," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3572–3585, Aug. 2008.
- [12] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: 1-D space," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3586–3599, Aug. 2008.
- [13] J. Kovacevic and M. Püschel, "Algebraic signal processing theory: Sampling for infinite and finite 1-D space," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 242–257, Jan. 2010.
- [14] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Graph filters," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 6163–6166.
- [15] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," in *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.
- [16] B. Steinberg, "Representation theory of finite groups: An introductory approach. Universitext. New York, NY, USA: Springer, 2011.
- [17] A. Terras, *Fourier Analysis on Finite Groups and Applications*, London Mathematical Society Student Texts. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [18] W. Fulton, W. F. J. Harris, and J. Harris, "Representation Theory: A First Course, Graduate Texts in Mathematics. New York, NY, USA: Springer, 1991.
- [19] L. Lovász, *Large Networks and Graph Limits*, American Mathematical Society colloquium publications. American Mathematical Society, 2012.
- [20] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 520–537, Jan.–Mar. 2020.
- [21] L. Ruiz, F. O. LuizChamon, and A. Ribeiro, "Graphon signal processing," 2020, *arXiv:2003.05030*.
- [22] A. Parada-Mayorga, L. Ruiz, and A. Ribeiro, "Graphon pooling in graph neural networks," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 860–864.
- [23] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro, "Convolutional neural network architectures for signals supported on graphs," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1034–1049, Feb. 2019.
- [24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, 2016.
- [25] F. Gama, J. Bruna, and A. Ribeiro, "Stability of graph scattering transforms," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 8038–8048.
- [26] G. B. Folland, *A Course in Abstract Harmonic Analysis*, Textbooks in Mathematics. Boca Raton, FL, USA: CRC Press, 2016.
- [27] A. Deitmar and S. Echterhoff, *Principles of Harmonic Analysis*. Universitext. Springer International Publishing, 2014.
- [28] P. I. Etingof *et al.*, *Introduction to Representation Theory*, Student Mathematical Library. American Mathematical Society, 2011.
- [29] M. Barot, *Introduction to the Representation Theory of Algebras*. Springer International Publishing, 2014.
- [30] M. Lorenz, *A Tour of Representation Theory*, Graduate Studies in Mathematics. American Mathematical Society, 2018.
- [31] J. B. Conway, *A Course in Functional Analysis*, *Graduate Texts in Mathematics*. New York, NY, USA: Springer, 1994.
- [32] C. D. Aliprantis, *An Invitation to Operator Theory*, Graduate Studies in Mathematics. American Mathematical Society, 2002.
- [33] Y. Benyamini and J. Lindenstrauss, *Geometric Nonlinear Functional Analysis*, Number v. 48, no. 1 in American Mathematical Society colloquium publications. American Mathematical Society, 2000.
- [34] J. Lindenstrauss, D. Preiss, and J. Tišer, *Frechet Differentiability of Lipschitz Functions and Porous Sets in Banach Spaces*, Annals of Mathematics Studies. Princeton, NJ, USA: Princeton Univ. Press, 2012.
- [35] M. S. Berger, *Nonlinearity and Functional Analysis: Lectures on Nonlinear Problems in Mathematical Analysis*, *Pure and Applied Mathematics*. Amsterdam, The Netherlands: Elsevier Science, 1977.
- [36] N. J. Higham, *Functions of Matrices: Theory and Computation*, Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics. Philadelphia, PA, USA: SIAM, 2008.