THE GEOMETRY OF POLYNOMIAL GROUP CONVOLUTIONAL NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

We prove that polynomial group convolutional neural networks (PGCNNs), for an arbitrary finite group G, are identifiable. In order to achieve this, we introduce a new mathematical framework for PGCNNs using the language of graded group algebras. This framework yields two natural parametrizations of the architecture, based on Hadamard and Kronecker products, related by a linear map. We compute the dimension of the associated neuromanifold, verifying that it depends only on the number of layers and the size of the group. The identifiability of PGCNNs follows from showing that the general fiber of both parametrizations is trivial up to the regular group action and rescaling.

1 INTRODUCTION

Geometric deep learning (GDL) (Bronstein et al., 2017; 2021; Gerken et al., 2023) is the subfield in deep learning that studies equivariant architectures with respect to a given group G. The main principle of GDL is when training over a dataset that has a set of symmetries, we can introduce an informed bias towards equi/invariant models by using equivariant architectures such as convolutional layers or graph layers. The benefit is a reduction of sample complexity; an equivariant architecture has fewer parameters due to weight-sharing, and fewer parameters need fewer training data points to determine their values and, henceforth, lower sample complexity (Du et al., 2018).

A classical example of equivariant architectures is convolutional neural networks (CNN), which are equivariant under discrete translations, corresponding to the group $G = \mathbb{Z}$. A general theory for linear equivariant layers for finite groups and compact groups is presented in Cohen & Welling (2016); Cohen et al. (2019). The core idea is that enforcing equivariance on linear maps, under given representation, results in weight-sharing constraints which G-convolutional layers have. In Nyholm et al. (2025), this approach is extended to non-linear equivariant layers.

Apart from designing equivariant models, another direction of geometric deep learning is concerned with understanding the expressivity and training dynamics of such models. In this article we are interested in the latter direction, and to investigate this, we apply tools from the recently emergent field of *neuroalgebraic geometry* (Marchetti et al., 2025).

The main idea of neuroalgebraic geometry can be summarized as follows. Given a fixed neural network architecture Φ, one obtains a parametric family of functions Φ_θ indexed by the network parameters θ. We call the space of all realizable functions *the neuromanifold* $\mathcal{M}_\Phi$ associated with the architecture Φ. The term "manifold" is a slight misnomer, as these functional spaces usually exhibit singularities. Given a loss function $\mathcal{L}$ and a dataset $\mathcal{D}$, then learning via stochastic gradient descent can be interpreted as a distance minimization problem over the neuromanifold to an objective function $f_{\mathcal{L},\mathcal{D}}$ in the ambient space, that depends on both the loss function $\mathcal{L}$ and the dataset $\mathcal{D}$. From this perspective, the geometric properties of the neuromanifold translate into statistical counterparts. For example, dimension and curvature translate into sample complexity and expressivity, fiber classification translates into identifiability, and so on.

Let us expand on the last point. An architecture is identifiable if the weights of hidden layers are determined by the relation between the inputs and outputs of the network, up to symmetries such as rescaling and neuron permutation (Sussmann, 1992; Albertini & Sontag, 1993; Fefferman, 1994). In other words, identifiability implies that the general fiber of the parametrization map is an orbit of the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

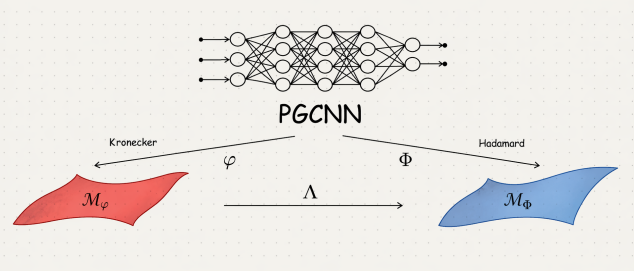

Figure 1: PGCNN parametrization maps φ and Φ with their images $\mathcal{M}_\varphi$ and $\mathcal{M}_\Phi$, related by the linear map Λ. $\mathcal{M}_\Phi$ is the neuromanifold of the PGCNN.

architecture's symmetries. This is also relevant for optimization because along the fiber directions any loss is flat (Trager et al., 2020).

The analysis of neuromanifolds is often intractable due to their compositional complexity and the use of nonlinear activation functions, which in some cases even fail to be analytic (e.g. the ReLU function). One useful simplification is to restrict to polynomial activation functions, which can be justified for three reasons. Firstly, polynomials are universal approximators by Weierstrass Approximation Theorem, which implies that neural networks with non-polynomial activation functions can be approximated by networks with polynomial activation functions. Secondly, the neuromanifold becomes a subspace of the space of homogeneous polynomials of fixed degree and in a fixed number of variables. More precisely, the neuromanifold becomes a semi-algebraic set (i.e. it can be described as a solution set of polynomial equations and inequalities in its finite dimensional ambient space), which can be analyzed using tools from algebraic geometry. Thirdly, although non-polynomial activations are more common in mainstream applications, neural networks with polynomial activation have performed well in many different applications (e.g healthcare, finance, computer vision, cryptography, theoretical physics) (Dey et al., 2016; Fong et al., 2022; Ghazali et al., 2011; Nayak & Misra, 2018; Yavartanoo et al., 2021; Chrysos et al., 2021; Hesamifard et al., 2017; Douglas et al., 2022).

Main contributions. The aim of this paper is to generalize the results in Shahverdi et al. (2025) from polynomial CNNs to polynomial group convolutional neural networks (PGCNN) for an arbitrary finite group G. We formalize the construction of PGCNNs using the language of group algebras and describe the neuromanifolds as subspaces of these algebras. In particular, we study the dimension of the neuromanifold which is related to expressivity and sample complexity. We also investigate the general fiber of the parametrization map which determines the identifiability of the architecture. Our results are verified numerically in Appendix I.

The main contributions of this paper are summarized as follows:

- We prove that PGCNNs are identifiable for every finite group G and every activation degree $r \geq 2$ (Theorem 4.5).
- Our approach describes the parametrization map Φ of a PGCNN through graded group algebras and encodes the monomial activation as a Hadamard product (§2 and §3).
- For $r \geq 2$ we compute the dimension of the neuromanifold $\mathcal{M}_\Phi$ and show that it depends only on the depth of the network and on the size $|G|$ of the group (Theorem 4.1).
- Consequently, a PGCNN and an equi-width PNN of width $|G|$ have the same ratio of parameter count to dimension. Dimension therefore cannot account for the greater efficiency of equivariant models; this must instead be reflected in the curvature of the neuromanifold.
- We factor the parametrization as $\Phi = \Lambda \circ \varphi$, where Λ is linear and φ is obtained from Φ by replacing every Hadamard product with a Kronecker product. Throughout the article we show that φ shares the properties we establish for Φ, dimension and identifiability included. The results for φ, however, follow from tensor linear algebra, whereas those for Φ require algebraic geometry. We read this as evidence that studying the tensorial part of polynomial networks might be enough.

Related works. We review here works related to neuroalgebraic geometry and focus on those concerning convolutional and equivariant architectures. Mehta et al. (2022) presented one of the first attempts to use algebraic geometry to count critical points in the loss landscape for deep linear neural networks. Building on this perspective, Trager et al. (2020) introduced the classification of critical points as spurious (due to parametrization) and pure (due to geometry). In (Kileel et al., 2019), the authors connected the dimension of the neuromanifold of PNNs to the expressivity of the model and established upper and lower bounds on layer width for neuromanifolds to be filling (i.e. they fill their ambient space). In the same paper, they conjectured that for sufficiently high activation degrees, all PNN architectures are filling. This conjecture was later proven independently in each of (Finkel et al., 2025; Massarenti & Mella, 2025; Craciun, 2025). Also, questions about non-defectiveness and identifiability of PNNs were settled separately by Usevich et al. (2025) using techniques from tensor decompositions and by Massarenti & Mella (2025) by analyzing secant varieties.

Turning to convolutional architectures, the neuromanifold of a linear convolutional network was first studied in (Kohn et al., 2022) and (Kohn et al., 2024). In these works, convolution filters were identified with homogeneous bivariate polynomials and the convolution operation with polynomial multiplication. These allowed the authors to define the neuromanifold by factorization conditions and to show it is a semi-algebraic set; moreover, conditions under which it is filling are established. Building on this line of work, Shahverdi (2024) developed recursive algorithms to construct the polynomial equations and inequalities that define the neural manifold for a linear convolutional network, and it relates the number of critical points of the square loss function to the generic Euclidean degree of the Segre variety.

The work most closely related to this article is Shahverdi et al. (2025). In this work, the authors compute the dimension, the degree and the generic Euclidean degree for polynomial convolutional networks in addition to classifying the general fiber and the singular locus. A central aspect of their analysis is based on the relation between the neuromanifold and the Segre-Veronese variety; in particular, they show that the neuromanifold for CNN is birational to the Segre-Veronese variety. Our work generalizes their dimension result for PGCNNs. In a later work (Shahverdi et al., 2026b), the authors considered neural networks and CNN for a generic polynomial as an activation functions, and they showed that for CNN the fibers are generally singletons and for fully connected neural networks fibers are finite. Finally, in Shahverdi et al. (2026a), they treated identifiable equivariant models and showed that these models are layerwise equivariant for some representation on latent spaces. An interesting feature of this result is that it is architecture-agnostic.

The results of Massarenti & Mella (2025); Usevich et al. (2025); Finkel et al. (2025); Craciun (2025) show that the neuromanifolds of polynomial neural networks (PNN) with fixed width have the expected (maximum) dimension. More precisely, for a *general element* of such a neuromanifold, the tangent space has the expected dimension. Since a PGCNN can be represented as a PNN with fixed width, then one might expect that the dimension of the neuromanifold of PGCNN can be inferred immediately from these results. This is not true because a PGCNN enforces weight-sharing constraints that violate the required generality assumption. In terms of algebraic geometry, the neuromanifold of PGCNN is a specialization of the neuromanifold of PNN with fixed width, and general conditions do not necessarily hold over specializations.

Outline. In Section 2, we present a concise introduction of graded group algebras. We then construct PGCNNs and their neuromanifolds in Section 3. Finally, in Section 4 we state all our main results regarding the dimension of neuromanifolds and the shape of general fibers of PGCNNs. We conclude and present ideas for future research in Section 5. Proofs are relegated to the appendix.

Notation. We denote by $\mathbb{K}$ a field with characteristic 0 and its closure by $\overline{\mathbb{K}}$. For example, $\mathbb{K}$ can be the real numbers, the complex numbers, the rational numbers or p-adic numbers. For a field $\mathbb{K}$ and n variables $x = (x_1, \dots, x_n)$ we denote the ring of polynomials in x over the field $\mathbb{K}$ by $\mathbb{K}[x_1, \dots, x_n]$ and the space of homogeneous polynomials in the variables x of degree d by $\mathrm{Sym}_{\mathbb{K}}(x, d)$. More generally, let the variables be partitioned into m blocks $x^{(1)}, \dots, x^{(m)}$, where $x^{(i)} = (x_1^{(i)}, \dots, x_{n_i}^{(i)})$. A polynomial $f \in \mathbb{K}[x^{(1)}, \dots, x^{(m)}]$ is called *multihomogeneous* of multidegree $(d_1, \dots, d_m)$ if it is homogeneous of degree d_i with respect to each block $x^{(i)}$. We denote the space of such polynomials by $\mathrm{Sym}_{\mathbb{K}}\big((x^{(1)}, \dots, x^{(m)}), (d_1, \dots, d_m)\big)$.

2 PRELIMINARIES ON GRADED GROUP ALGEBRAS

In this section, we present graded group algebras, their regular representations, and how Kronecker and Hadamard products act on them. These are the building blocks for PGCNNs. Readers familiar with these concepts may skip this section and return to it when needed.

In the following, G and H are general finite groups with order $n = |G|$ and $|H|$ and $S := \bigoplus_i S_i$ is a graded commutative ring such as the ring of polynomials in n variables $\mathbb{K}[x_1, \dots, x_n]$. For two elements $g, h \in G$ we denote their group product by gh and for two elements $s, t \in S$ we denote their ring product by st and their sum by $s + t$.

Definition 2.1. The graded group algebra $S[G]$ is an algebra $S[G] := \{\theta \mid \theta : G \to S\} \cong S^{|G|}$, where $S^{|G|}$ is the free S-module of rank $|G|$. The graded algebra bilinear product in $S[G]$ is given by *the convolution operation,*

$$(\theta * \psi)(g) := \sum_{h \in G} \theta(gh^{-1})\psi(h),$$

and the scalar multiplication is defined componentwise; i.e. for $s \in S$ we define $(s\theta)(g) := s\theta(g)$.

If H is a subgroup of G then $S[H]$ is a subalgebra of $S[G]$. Therefore, for $\psi \in S[H]$ and $\theta \in S[G]$ we have $\theta * \psi, \psi * \theta \in S[G]$. Namely, for $g \in G$ we define

$$(\theta * \psi)(g) := \sum_{h \in H} \theta(gh^{-1})\psi(h) \text{ and } (\psi * \theta)(g) := \sum_{h \in H} \psi(h)\theta(h^{-1}g).$$

In other words, any filter $\psi \in S[H]$ can be extended to G by setting $\psi(g) = 0$ for all $g \notin H$. If we need to distinguish between $\psi \in S[H]$ and its extension $\psi \in S[G]$ we denote the extension by ψ^G.

Definition 2.2 ($S[G]$ regular representation)**.** The regular algebra representation ρ of $S[G]$ on itself is defined as $\rho : S[G] \to \text{End}(S[G]), \theta \mapsto \rho_\theta$ where $\rho_\theta(\psi) := \theta * \psi$.

Note that the algebra representation ρ is completely determined by the induced group representation of G on $S[G]$. Moreover, if an order $(g_1, \dots, g_n)$ of the elements of G is fixed then, with respect to this basis of $S[G]$, the map ρ_θ is represented by an $n \times n$ *G-circulent matrix* Mat_θ over S, whose entries are given by $(\text{Mat}_\theta)_{ij} := \theta(g_i g_j^{-1})$. In the special case $G = \mathbb{Z}_n$, the matrix Mat_θ reduces to the classical circulent matrix.

Next we define the Kronecker product and the Hadamard product between filters and state propositions about their interactions with convolution. These two operations will give us two ways to define polynomial activation functions.

Proposition 2.3 (Definition of Kronecker Product)**.** *There is a canonical isomorphism ι between $S[G] \otimes S[H]$ and $S[G \times H]$ given by*

$$\iota : S[G] \otimes S[H] \to S[G \times H]$$
$$\theta \otimes \psi \mapsto \iota(\theta \otimes \psi)$$

where for $(g, h) \in G \times H$, $\iota(\theta \otimes \psi)(g, h) := \theta(g)\psi(h)$. By abuse of notation, we denote the filter $\iota(\theta \otimes \psi)$ by $\theta \otimes \psi$ and call it the Kronecker product of θ and ψ.

Remark 2.4. For $r \geq 1$, and $\theta \in S[G]$ we use the notation $\theta^{\otimes r} := \underbrace{\theta \otimes \dots \otimes \theta}_{r}$, which is a filter in $S[G^r]$ where G^r is the direct product of r copies of G. Note that the Kronecker product respects the grading; i.e. if $\theta \in S_k[G]$ and $\psi \in S_l[H]$ then $\theta \otimes \psi \in S_{k+l}[G \times H]$ and $\theta^{\otimes r} \in S_{rk}[G^r]$.

Definition 2.5 (Hadamard Product)**.** Let $\theta, \psi \in S[G]$, their *Hadamard product* is the filter denoted by $\theta \odot \psi \in S[G]$ and is given by $(\theta \odot \psi)(g) := \theta(g)\psi(g)$. Moreover, we denote the rth Hadamard power of θ by $\sigma_r(\theta) := \underbrace{\theta \odot \dots \odot \theta}_{r}$. The Hadamard product also respects the grading of S.

To justify our terminology, note that for any two filters θ and ψ we have the identities $\text{Mat}_{\theta \otimes \psi} = \text{Mat}_\theta \otimes \text{Mat}_\psi$ and $\text{Mat}_{\theta \odot \psi} = \text{Mat}_\theta \odot \text{Mat}_\psi$ where on the right hand side in both of them, we use the Kronecker product and the Hadamard product between matrices, respectively (See Appendix A for concrete examples on small groups).

The following proposition shows that the Kronecker product distributes over convolution.

Proposition 2.6. *Let $\theta_1, \theta_2 \in S[G]$ and $\psi_1, \psi_2 \in S[H]$. Then $(\theta_1 \otimes \psi_1) * (\theta_2 \otimes \psi_2) = (\theta_1 * \theta_2) \otimes (\psi_1 * \psi_2)$. By induction it holds that for $\theta_i, \psi_i \in S[G_i]$ for $1 \leq i \leq r$*

$$(\theta_1 \otimes \ldots \otimes \theta_r) * (\psi_1 \otimes \ldots \otimes \psi_r) = (\theta_1 * \psi_1) \otimes \ldots \otimes (\theta_r * \psi_r).$$

On the other hand, the Hadamard product does not distribute over convolution, and this is easy to check as follows. Let $\theta_1, \theta_2, \psi_1, \psi_2 \in S[G]$ then

$$\begin{aligned}(\theta_1 \odot \theta_2) * (\psi_1 \odot \psi_2)(g) &= \sum_{h \in G} \theta_1(gh^{-1})\theta_2(gh^{-1})\psi_1(h)\psi_2(h) \\ &\neq \left(\sum_{h \in G} \theta_1(gh^{-1})\psi_1(h)\right)\left(\sum_{h \in G} \theta_2(gh^{-1})\psi_2(h)\right) = (\theta_1 * \psi_1) \odot (\theta_2 * \psi_2)(g).\end{aligned}$$

The following proposition explains the relation between the Hadamard product, the Kronecker product, and convolution. This is an important proposition, as it tells us how to move between the two descriptions of polynomial activation functions; see §3.

Proposition 2.7. *Let $\theta_1, \psi_1, \theta_2, \psi_2 \in S[G]$. Then*

$$(\theta_1 * \psi_1) \odot (\theta_2 * \psi_2) = ((\theta_1 \otimes \theta_2) * (\psi_1 \otimes \psi_2))\,|_G.$$

It follows that for $\theta, \psi \in S[G]$ and $r \geq 1$ we get $\sigma_r(\theta * \psi) = (\theta^{\otimes r} * \psi^{\otimes r})|_G$, where G is identified with the diagonal subgroup in G^r.

3 Polynomial group convolutional neural networks

In this section, polynomial group convolutional networks and their neuromanifolds are defined, which are the main object of study here. Recall that a main goal of this article is to compute the dimension of these neuromanifolds.

3.1 Definition of polynomial group convolutional neural networks

Definition 3.1 (GCNN and PGCNN)**.** A group convolutional neural network (GCNN) with the filters $\theta := (\theta_1, \ldots, \theta_L) \in (\mathbb{K}[G])^L$ and an activation function $\sigma : \mathbb{K} \to \mathbb{K}$ is the map

$$\begin{aligned}\Phi_\theta : \mathbb{K}[G] &\to \mathbb{K}[G] \\ x &\mapsto \sigma(\ldots(\sigma(\sigma(x * \theta_1) * \theta_2) * \ldots * \theta_{L-1}) * \theta_L,\end{aligned}$$

where σ acts component wise. When $\sigma = \sigma_r$, where σ_r denotes the rth Hadamard product, we call Φ_θ a polynomial group convolutional neural network (PGCNN). We call x the input filter and θ_l the lth weight filter for $1 \leq l \leq L$.

The GCNNs are G-equivariant, where the G-action is the one given by the regular representation in Section §2, and equivariance follows directly from the associativity of convolution. Before discussing PGCNN further, we present a few remarks to connect our definition with existing work in the literature.

Remark 3.2. Convolutional neural networks, regardless of their name, are in the literature defined using the operation *cross correlation* $\star$ instead of convolution $*$. Nevertheless, these two operations are intrinsically connected and interchangeable via the relation $\theta \star \psi := \theta * \overline{\psi}$, where the filter $\overline{\psi}(g) := \psi(g^{-1})$ is called the involution of ψ. Therefore these two conventions are equivalent in the sense that every neural network defined via convolution can be defined via cross correlation by involuting all its filters. In this article, we adopt the convolutional convention for two reasons. Firstly, under this convention, the G-action translates to just one extra left convolution, and the equivariance property translates into associativity of the convolution operation. Secondly, by associativity of convolutions, we can study the interaction between the weight filters θ away from the input filter x.

Remark 3.3. In machine learning literature, the input filter x is called the input signal, and the weight filters correspond to convolutional layers. Moreover, the signal x is considered over a domain Ω equipped with a G action. In other words, $x : \Omega \to \mathbb{K}$ where for $t \in \Omega$ and $g \in G$ we have $(g \cdot x)(t) := x(g^{-1} \cdot t)$. In this context, what we define above is a special case of GCNN where the domain $\Omega = G$. This simplification is convenient because it keeps the group algebra structure at the first layer.

Remark 3.4. Notice that when $G = \mathbb{Z}_n$ is the nth cyclic group, the GCNN does *not* specialize to CNN considered, for example, in Kohn et al. (2022) and Shahverdi et al. (2025) because finite-filter CNNs on finite sequences, without circular padding, are *not* equivariant with respect to cyclic shifts. The main reason is that in this CNN architecture, the domain is considered to be a sequence rather than a circle. To see this, let C_k denotes the domain that is a sequence of length k. Moreover, let x be a signal on C_n and θ be a signal (or filter) over C_k where $k < n$. Then their cross correlation with a step of length 1 will be the signal $x \star_{CNN} \theta$ over C_{n-k+1} given by

$$(x \star_{CNN} \theta)(l) := \sum_{j=0}^{k-1} x(l+j)\theta(j), \text{ where } l < n-k+1.$$

Assume that s_d denotes the shift to the right with d steps; then $(s_d \cdot x)(l) := x(l-d \mod n)$, note importantly that the order of the modulo operation depends on the size of the signal x. Given this, we have

$$s_d \cdot (x \star_{CNN} \theta)(l) = (x \star_{CNN} \theta)(l-d \mod (n-k+1)) = \sum_{j=0}^{k-1} x([l-d \mod (n-k+1)]+j)\theta(j),$$

and

$$(s_d \cdot x \star_{CNN} \theta)(l) = \sum_{j=0}^{k-1} s_d \cdot x(j+l)\theta(j) = \sum_{j=0}^{k-1} x([l-d \mod n]+j)\theta(j).$$

This implies $s_d \cdot (x \star_{CNN} \theta) \neq s_d \cdot x \star_{CNN} \theta$, unless $k = 1$ which means the convolution filter contains only one element. Hence, standard discrete CNNs with finite filters are not equivariant with respect to cyclic shifts, except in the trivial case $k = 1$.

3.2 Neuromanifolds of PGCNNs

From Definition 3.1, we see that for any tuple of L filters $\theta := (\theta_1, \ldots, \theta_L) \in (\mathbb{K}[G])^L$ and activation function σ_r, the PGCNN $\Phi_\theta \in \mathrm{Sym}_{\mathbb{K}}(x, r^{L-1})[G]$ defines n homogeneous polynomials in n variables $x := (x_1, \ldots, x_n)$ with degree r^{L-1}. Moreover, each coefficient in the homogeneous polynomial $\Phi_\theta(g)$ is itself a multihomogeneous polynomial with the multidegree $\mathbf{r} := (r^{L-1}, r^{L-2}, \ldots, 1)$ in the parameters of the weight filters $\theta = (\theta_1, \ldots, \theta_L)$. Thus, these coefficients belong to the space $\mathrm{Sym}_{\mathbb{K}}(\theta, \mathbf{r})$.

Definition 3.5 (Neuromanifold)**.** The *neuromanifold* $\mathcal{M}_\Phi$ of the architecture of PGCNN of depth L and activation degree r is the image of the parameterization map $\Phi : \theta \mapsto \Phi_\theta$ in $\mathrm{Sym}_{\mathbb{K}}(x, r^{L-1})$; namely

$$\mathcal{M}_\Phi = \{\Phi_\theta \mid \theta \in (\mathbb{K}[G])^L\} \subset \mathrm{Sym}_{\mathbb{K}}(x, r^{L-1})[G].$$

Note that the projection map $\Phi_\theta \mapsto \Phi_\theta(e)$ over the identity element is injective due to equivariance. Therefore we can consider the ambient space of the neuromanifold $\mathcal{M}_\Phi$ to be $\mathrm{Sym}_{\mathbb{K}}(x, r^{L-1})$ rather than $\mathrm{Sym}_{\mathbb{K}}(x, r^{L-1})[G]$.

As we have seen until now, the activation function is represented using the Hadamard product. Another equivalent way is to replace the Hadamard product with the Kronecker product using Proposition 2.7. So in terms of the Kronecker product, the filter Φ_θ can be written as

$$\Phi_\theta = (x^{\otimes r^{L-1}} * \varphi_\theta)|_G,$$

where $\varphi_\theta = \theta_1^{\otimes r^{L-1}} * \theta_2^{\otimes r^{L-2}} * \ldots * \theta_L \in \mathbb{K}[G^{r^{L-1}}]$. We define the parametrization map

$$\begin{aligned} \varphi : \mathbb{K}[G]^L &\to \mathbb{K}[G^{r^{L-1}}] \\ \theta &\mapsto \varphi_\theta \end{aligned}$$

and we denote its image by $\mathcal{M}_\varphi := \{\varphi_\theta \mid \theta \in \mathbb{K}[G]^L\} \subset \mathbb{K}[G^{r^{L-1}}]$.

Both of the maps Φ and φ are defined by multihomogeneous polynomials in the weight parameters $\theta = (\theta_1, \ldots, \theta_L)$. Therefore, when $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, Tarski-Seidenberg Theorem and Chevalley Theorem, respectively imply that both $\mathcal{M}_\Phi$ and $\mathcal{M}_\varphi$ are semi-algebraic sets in their respective ambient

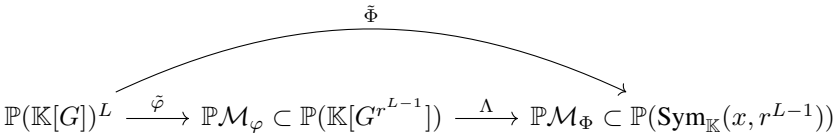

Figure 2: A commutative diagram that describes the relation between the projectivized parametrization maps $\tilde{\Phi}$ and $\tilde{\varphi}$ induced by the PGCNN architecture with L layers and activation degree r over the finite group G.

spaces. Moreover, they induce rational maps $\tilde{\Phi}$ and $\tilde{\varphi}$ on the projectivizations of their domains and codomains. The relation between the rational maps $\tilde{\varphi}$ and $\tilde{\Phi}$ is shown in a commutative diagram, see Figure 2, where Λ is a projective linear map (it consists of convolution with $x^{\otimes r^{L-1}}$ and the restriction to $e \in G$).

Remark 3.6 (Linear GCNN)**.** In the linear case, where $r = 1$, we have

$$\Phi_\theta = x * \varphi_\theta = x * \theta_1 * \ldots * \theta_L.$$

Note then that the coefficients of the polynomial $\Phi_\theta(e) \in \mathrm{Sym}_{\mathbb{K}}(x, 1)$ determine completely the filter $\varphi_\theta \in \mathbb{K}[G]$, and vice versa. Moreover, it is clear that φ is surjective. Therefore in the linear case, we have $\mathcal{M}_\Phi \cong \mathcal{M}_\varphi = \mathbb{K}[G]$.

4 Main results on neuromanifolds of PGCNN

In this section, we state our main results about the neuromanifolds of PGCNN: we compute the dimensions of $\mathcal{M}_\varphi$ and $\mathcal{M}_\Phi$, and give a description of the general fibers of φ and Φ. Recall that the interest in the dimension of the neuromanifold and the general fibers is because these geometrical concepts are tied to the sample complexity and the identifiability of the models (Marchetti et al., 2025).

4.1 Dimension of $\mathcal{M}_\varphi$ and $\mathcal{M}_\Phi$

Our technique to compute the dimension of the neuromanifold $\mathcal{M}_\Phi$ and the space $\mathcal{M}_\varphi$ decomposes broadly to

1. We study the rank of the Jacobians of the parametrization maps Φ and φ at general filters. Note that for nonzero scalars $\lambda_l \in \mathbb{K} - \{0\}$, where $1 \leq l \leq L-1$, and any tuple of filters $\theta = (\theta_1, \ldots, \theta_L) \in \mathbb{K}[G]^L$, we have
$$\varphi_\theta = \varphi_\psi \text{ and } \Phi_\theta = \Phi_\psi, \tag{1}$$
where $\psi = (\lambda_1\theta_1, \lambda_1^{-r}\lambda_2\theta_2, \ldots, \lambda_{L-2}^{-r}\lambda_{L-1}\theta_{L-1}, \lambda_{L-1}^{-r}\theta_L)$. Therefore the rank of the Jacobians $J_\theta\varphi$ and $J_\theta\Phi$ is bounded from above by $L(n-1)+1$.
2. The vectors that lie in the kernels of $J_\theta\varphi$ and $J_\theta\Phi$ due to Eq. (1) are all killed by the projectivization of the domain. Therefore if $J_\theta\tilde{\varphi}$ and $J_\theta\tilde{\Phi}$ are full rank, then the affine Jacobians $J_\theta\varphi$ and $J_\theta\Phi$ have the maximal rank $L(n-1)+1$.
3. To compute the rank of the Jacobians at general filters θ, we consider the Jacobians as symbolic matrices in terms of the parameters θ and compute their symbolic rank.

Following this technique, we prove the following theorem (See Appendices E and F for proofs).

Theorem 4.1. *For activation degree $r \geq 2$, the rational maps $\tilde{\varphi}$ and $\tilde{\Phi}$ are* generally regular, *i.e. the Jacobian of each map is full rank at a general point in its domain. This implies that* $\dim(\mathcal{M}_\varphi) = \dim(\mathcal{M}_\Phi) = L(n-1)+1$*, where $n = |G|$ and L is the number of layers.*

Fix an activation degree $r \geq 2$. For an equi-width PNN with width n and length L, the dimension of the neuromanifold is $n(L(n-1)+1)$ (Finkel et al., 2025; Usevich et al., 2025), which means the ratio

of dimension per parameter is the same for this PNN as the PGCNN with depth L and a finite group G with size equal to n. Consequently, we see that a better efficieny of equivariant/convolutional models should be reflected by the curvature of the neuromanifold which corresponds to the degree of the neurovariety (Zariski closure of the neuromanifold), instead of the dimension.

4.2 General fiber of φ

We study the general fiber of φ. For filters $\theta := (\theta_1, \dots, \theta_L)$, the fiber at φ_θ is $\varphi^{-1}(\varphi_\theta) = \{\psi \in \mathbb{K}[G]^L \mid \varphi_\psi = \varphi_\theta\}$. To describe the general fiber we use a standard fact on symmetric tensor rank.

Lemma 4.2 (Lemma 5.1 (Comon et al., 2008))**.** *If $y_1, \dots, y_n \in \mathbb{K}^d$ are linearly independent and $r \geq 2$, then the symmetric rank of the symmetric tensor $A = \sum_{i=1}^n y_i^{\otimes r}$ is equal to n.*

Theorem 4.3. *For activation degree $r \geq 2$, and general filters $\theta := (\theta_1, \dots, \theta_L)$, we have $\tilde{\varphi}_\theta = \tilde{\varphi}_\psi$ if and only if $\psi = (\theta_1 * g_1,\ g_1^{-1} * \theta_2 * g_2,\ \dots,\ g_{L-2}^{-1} * \theta_{L-1} * g_{L-1},\ g_{L-1}^{-1} * \theta_L)$, where $g_i \in G$ for $1 \leq i \leq L-1$, up to rescaling each filter.*

Proof sketch. Induct on L. Writing $\theta' = (\theta_1, \dots, \theta_{L-1})$ and $\psi' = (\psi_1, \dots, \psi_{L-1})$, the identity $\varphi_\theta = \lambda \varphi_\psi$ implies $\varphi_{\theta'}^{\otimes r} * \theta_L = \lambda\, \varphi_{\psi'}^{\otimes r} * \psi_L$. For general θ, the last layer ψ_L is invertible (Proposition D.5), hence with $\mu := \theta_L * \psi_L^{-1}$ we obtain $\varphi_{\theta'}^{\otimes r} * \mu = \lambda\, \varphi_{\psi'}^{\otimes r}$. Distributing convolution over $\otimes$ yields $\sum_{g \in G} \mu(g)\, (\varphi_{\theta'} * g)^{\otimes r} = \lambda\, \varphi_{\psi'}^{\otimes r}$. For general θ', the columns $\{\varphi_{\theta'} * g\}_{g \in G}$ are linearly independent (Proposition D.5), so Lemma 4.2 forces μ to be a (scaled) group element. Thus $\theta_L = g * \psi_L$ up to scale, and the inductive hypothesis gives the stated form for all layers. Full details are in Appendix H. □

4.3 General fiber of Φ

In this subsection, we discuss the general fiber of the parametrization map Φ. To do that, we develop a proposition similar in nature to Lemma 4.2, and by using it we are able to show that the general fiber Φ is the same as the general fiber of φ.

Denote by $VP_{n,k,d}$ the projective variety of kth power of degree d homogeneous polynomials in n variables x (Abdesselam & Chipalkatti, 2014). Namely,

$$VP_{n,k,d} := \{[p] \in \mathbb{P}(\mathrm{Sym}_{\mathbb{K}}(x, kd)) \mid \exists q \in \mathrm{Sym}_{\mathbb{K}}(x, d),\ q^k = p\}.$$

Proposition 4.4. *Let $G = \{g_1, \dots, g_n\}$ as a set, and let the activation degree $r \geq 2$. For general filters $\theta = (\theta_1, \dots, \theta_L)$, where $L \geq 1$, and $p_i(\theta) := \sigma_r(\Phi_\theta)(g_i)$ we have*

$$\mathbb{P}(\langle p_1(\theta), \dots, p_n(\theta)\rangle) \cap\ VP_{n,r,r^{L-1}} = \{[p_1(\theta)], \dots, [p_n(\theta)]\},, \tag{2}$$

where $\langle p_1(\theta), \dots, p_n(\theta)\rangle$ denotes the linear span of these polynomials in $\mathrm{Sym}_{\mathbb{K}}(x, r^L)$.

Proof sketch. We consider the incidence scheme

$$Y := \{(\theta, [F]) \in \Theta^o \times VP_{n,r,r^{L-1}} \mid p_1(\theta) \wedge \cdots \wedge p_n(\theta) \wedge F = 0 \in \bigwedge^{n+1} \mathrm{Sym}_{\mathbb{K}}(x, r^L)\},$$

where $\Theta^o \subset \Theta$ is the open subspace of the parameters such that $\{p_1(\theta), \dots, p_n(\theta)\}$ are linearly independent. We consider the projection map $\pi : Y \to \Theta^o$ and note that for $\theta \in \Theta^o$ the fiber $\pi^{-1}(\theta)$ is the intersection defined by (2). Using two different semi continuity theorems from algebraic geometry we restrict the dimension of fiber to be 0 for a general element $\theta \in \Theta^o$ and then we restrict its size (length) to be equal to n.

Note that the first part of the argument, where we restrict the dimension of the fiber, is similar to an argument in Finkel et al. (2025, Theorem 22) in which they compute the dimension of the neurovariety for a fully connected neural networks with layers of equal width. Our argument extends this to show the exact size of the general fiber. See Appendix G for the complete proof. □

Theorem 4.5. *For activation degree $r \geq 2$ and general filters $\theta := (\theta_1, \dots, \theta_L)$, we have $\tilde{\Phi}_\theta = \tilde{\Phi}_\psi$ if and only if $\psi := (\theta_1 * g_1,\ g_1^{-1} * \theta_2 * g_2,\ \dots,\ g_{L-2}^{-1} * \theta_{L-1} * g_{L-1},\ g_{L-1}^{-1} * \theta_L)$, where $g_i \in G$ for $1 \leq i \leq L-1$, up to rescaling each filter.*

Proof sketch. Induct on L as before. From $\Phi_\theta = \lambda\Phi_\psi$ we get $\sigma_r(\Phi_{\theta'}) * \theta_L = \lambda\,\sigma_r(\Phi_{\psi'}) * \psi_L$, where $\theta' := (\theta_1, \dots, \theta_{L-1})$. For general θ, invertibility of the last layer implies that with $\sigma_r(\Phi_{\theta'}) * \mu = \lambda\,\sigma_r(\Phi_{\psi'})$, where $\mu := \theta_L * \psi_L^{-1}$. Evaluating at $e \in G$ shows $\lambda\sigma_r(\Phi_{\psi'})(e)$ lies in the span $\langle\sigma_r(\Phi_{\theta'})(g)\rangle_{g\in G}$. Proposition 4.4 forces $\lambda\sigma_r(\Phi_{\psi'})(e) = \sigma_r(\Phi_{\theta'})(g)$ for some $g \in G$, hence μ is a scaled group element and $\theta_L = \lambda\, g * \psi_L$. Conclude by the inductive hypothesis. Full details are in Appendix H. □

According to our definition of identifiability, Theorems 4.3 and 4.5 imply that φ and Φ are identifiable, which corresponds to *globally identifiable* in the sense of Usevich et al. (2025).

Remark 4.6. In Finkel et al. (2025), it is shown that the powers of non-proportional multivariate polynomials are linearly independent for all powers $r > \tilde{r}(k)$ where $\tilde{r}$ is a function that depends only the number of polynomials k. This is then used to find an activation degree threshold for an arbitrary architecture of fully connected networks. A similar line of reasoning can be followed to prove weaker version Proposition 4.4, where $r > \tilde{r}(|G|)$. However, in order to prove Proposition 4.4 for $r \geq 2$, we need to resort to the use of semi continuity theorems of fiber dimension and fiber length from algebraic geometry.

5 Conclusion and Outlook

In this work we proved that polynomial group convolutional networks are identifiable for every finite group G and every activation degree $r \geq 2$, by describing the general fibers of the parametrization map Φ. Our approach rests on a mathematical framework for PGCNNs based on graded group algebras, in which the monomial activation is encoded as a Hadamard product. Within this framework we computed the dimension of the neuromanifold of a PGCNN and showed that it depends only on $|G|$ and on the number of layers. A PGCNN therefore has the same ratio of parameters to dimension as an equi-width PNN of width $|G|$, so any efficiency gain of equivariant and convolutional models must be reflected in the curvature of the neuromanifold or the degree of the neurovariety rather than in its dimension. Alongside these results we carried out the same computations for the map φ obtained from Φ by replacing every Hadamard product with a Kronecker product, so that $\Phi = \Lambda \circ \varphi$ with Λ linear. Both the dimension and the identifiability statements hold for φ as well, but there they follow from tensor linear algebra alone, whereas Φ requires algebraic geometry. This suggests that checking such properties on the tensorial part of a polynomial network might be enough.

For future work, we aim to compute the degree and the generic Euclidean degree for the neuromanifold of a PGCNN. In the CNN case, considered in Shahverdi et al. (2025), the parametrization map Φ factors through Segre-Veronese embedding via a projective linear map. Since the degree of a variety is invariant under projective linear morphism, the authors get that the degree of the Segre-Veronese variety naturally translates into the degree of the neurovariety of a CNN. In contrast, in the case of PGCNN, the parametrization map Φ factors through the Segre-Veronese embedding via a rational map with a nontrivial kernel rather than a projective linear morphism. Therefore, in the case of PGCNN, we expect the Segre-Veronese factorization of the map Φ to give an upper bound for both the degree and the generic Euclidean degree, rather than determining them exactly.

For further generalization, we propose two natural directions. The first direction is to consider polynomial activation functions which are sums of Hadamard products. This extension is interesting since, as we discussed before, PNNs are being used in multiple fields now. Moreover, by Weierstrass Approximation Theorem, these networks can be used to approximate neural networks with non-polynomial activation functions. This might enable a deeper investigation of the geometric structure underlying GCNNs with non-polynomial activations.

Another direction is to relax our restriction to finite groups, and study PGCNNs of compact groups. In this case input signals and convolution filters are local sections of an associated vector bundle over a homogeneous space, and the convolution operation is formulated as an integral (Cohen et al., 2019). To consider this case, we either need to generalize our framework to use infinite dimensional graded group algebras, or to discretize compact groups to finite groups and reduce to the current case. While in this setting one would have an infinite-dimensional generalization of a neuromanifold, relevant questions still remain about fibers and singularities.

REFERENCES

Abdelmalek Abdesselam and Jaydeep Chipalkatti. On Hilbert Covariants. *Canadian Journal of Mathematics*, 66(1):3–30, February 2014. ISSN 1496-4279. doi: 10.4153/cjm-2012-046-1. URL http://dx.doi.org/10.4153/CJM-2012-046-1.

Francesca Albertini and Eduardo D. Sontag. For neural networks, function determines form. *Neural Networks*, 6(7):975–990, 1993. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(09)80007-5. URL https://www.sciencedirect.com/science/article/pii/S0893608009800075.

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42, 2017. doi: 10.1109/MSP.2017.2693418.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021. URL https://arxiv.org/abs/2104.13478.

Grigorios Chrysos, Stylianos Moschoglou, Giorgos Bouritsas, Jiankang Deng, Yannis Panagakis, and Stefanos Zafeiriou. Deep polynomial neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4021–4034, 2021. doi: 10.1109/TPAMI.2021.3058891.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

Taco S. Cohen, Mario Geiger, and Maurice Weiler. *A general theory of equivariant CNNs on homogeneous spaces*. Curran Associates Inc., Red Hook, NY, USA, 2019.

Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008. doi: 10.1137/060661569. URL https://doi.org/10.1137/060661569.

Alexandru Craciun. Linear independence of powers for polynomials, 2025. URL https://arxiv.org/abs/2507.10163. arXiv:2507.10163.

S. Dey, S. Naskar, T. Mukhopadhyay, U. Gohs, A. Spickenheuer, L. Bittrich, S. Sriramula, S. Adhikari, and G. Heinrich. Uncertain natural frequency analysis of composite plates including effect of noise – a polynomial neural network approach. *Composite Structures*, 143:130–142, 2016. ISSN 0263-8223. doi: https://doi.org/10.1016/j.compstruct.2016.02.007. URL https://www.sciencedirect.com/science/article/pii/S0263822316300277.

Michael Douglas, Subramanian Lakshminarasimhan, and Yidi Qi. Numerical calabi-yau metrics from holomorphic networks. In Joan Bruna, Jan Hesthaven, and Lenka Zdeborova (eds.), *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*, volume 145 of *Proceedings of Machine Learning Research*, pp. 223–252. PMLR, 16–19 Aug 2022. URL https://proceedings.mlr.press/v145/douglas22a.html.

Simon S Du, Yining Wang, Xiyu Zhai, Sivaraman Balakrishnan, Russ R Salakhutdinov, and Aarti Singh. How many samples are needed to estimate a convolutional neural network? In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/03c6b06952c750899bb03d998e631860-Paper.pdf.

Charles Fefferman. Reconstructing a neural net from its output. *Revista Matemática Iberoamericana*, 10(3):507–555, 1994. doi: 10.4171/rmi/160.

Bella Finkel, Jose Israel Rodriguez, Chenxi Wu, and Thomas Yahl. Activation degree thresholds and expressiveness of polynomial neural networks. *Algebraic Statistics*, 16(2):113–130, 2025. doi: 10.2140/astat.2025.16.113. URL https://msp.org/astat/2025/16-2/p02.xhtml.

Simon James Fong, João Alexandre Lobo Marques, G. Li, N. Dey, Rubén G. Crespo, E. Herrera-Viedma, F. Nauber Bernardo Gois, and José Xavier Neto. *The Comparison of Different Linear and Nonlinear Models Using Preliminary Data to Efficiently Analyze the COVID-19 Outbreak*, pp. 65–81. Springer International Publishing, Cham, 2022. ISBN 978-3-030-95281-5. doi: 10.1007/978-3-030-95281-5_3. URL https://doi.org/10.1007/978-3-030-95281-5_3.

Jan E. Gerken, Jimmy Aronsson, Oscar Carlsson, Hampus Linander, Fredrik Ohlsson, Christoffer Petersson, and Daniel Persson. Geometric deep learning and equivariant neural networks. *Artificial Intelligence Review*, 56(12):14605–14662, 2023. doi: 10.1007/s10462-023-10502-7. URL https://doi.org/10.1007/s10462-023-10502-7.

Rozaida Ghazali, Abir Jaafar Hussain, and Panos Liatsis. Dynamic ridge polynomial neural network: Forecasting the univariate non-stationary and stationary trading signals. *Expert Systems with Applications*, 38(4):3765–3776, 2011. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2010.09.037. URL https://www.sciencedirect.com/science/article/pii/S0957417410009942.

Ehsan Hesamifard, Hassan Takabi, and Mehdi Ghasemi. Cryptodl: Deep neural networks over encrypted data. *CoRR*, abs/1711.05189, 2017. URL http://arxiv.org/abs/1711.05189.

Joe Kileel, Matthew Trager, and Joan Bruna. On the expressive power of deep polynomial neural networks. *CoRR*, abs/1905.12207, 2019. URL http://arxiv.org/abs/1905.12207.

Kathlén Kohn, Thomas Merkh, Guido Montúfar, and Matthew Trager. Geometry of linear convolutional networks. *SIAM Journal on Applied Algebra and Geometry*, 6(3):368–406, 2022. doi: 10.1137/21M1441183. URL https://doi.org/10.1137/21M1441183.

Kathlén Kohn, Guido Montúfar, Vahid Shahverdi, and Matthew Trager. Function space and critical points of linear convolutional networks. *SIAM Journal on Applied Algebra and Geometry*, 8(2):333–362, 2024. doi: 10.1137/23M1565504. URL https://doi.org/10.1137/23M1565504.

Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, and Kathlén Kohn. Position: Algebra unveils deep learning - an invitation to neuroalgebraic geometry. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025. URL https://openreview.net/forum?id=mzc1KPkIMJ.

A. Massarenti and M. Mella. The alexander-hirschowitz theorem for neurovarieties, 2025. URL https://arxiv.org/abs/2511.19703.

Dhagash Mehta, Tianran Chen, Tingting Tang, and Jonathan D. Hauenstein. The loss surface of deep linear networks viewed through the algebraic geometry lens. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5664–5680, 2022. doi: 10.1109/TPAMI.2021.3071289.

Sarat Chandra Nayak and Bijan Bihari Misra. Estimating stock closing indices using a ga-weighted condensed polynomial neural network. *Financial Innovation*, 4(1):21, 2018. ISSN 2199-4730. doi: 10.1186/s40854-018-0104-2. URL https://doi.org/10.1186/s40854-018-0104-2.

Elias Nyholm, Oscar Carlsson, Maurice Weiler, and Daniel Persson. Equivariant non-linear maps for neural networks on homogeneous spaces, 2025. URL https://arxiv.org/abs/2504.20974.

Vahid Shahverdi. Algebraic complexity and neurovariety of linear convolutional networks. *CoRR*, abs/2401.16613, 2024. URL https://doi.org/10.48550/arXiv.2401.16613.

Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. On the geometry and optimization of polynomial convolutional networks. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. URL https://openreview.net/forum?id=FdYA56Gcsj.

Vahid Shahverdi, Giovanni Luca Marchetti, Georg Bökman, and Kathlén Kohn. Identifiable equivariant networks are layerwise equivariant, 2026a. URL https://arxiv.org/abs/2601.21645.

Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. Learning on a razor's edge: Identifiability and singularity of polynomial neural networks. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=L5jYWeycAx.

The Stacks project authors. The stacks project. https://stacks.math.columbia.edu, 2026.

Héctor J. Sussmann. Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks*, 5(4):589–593, 1992. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(05)80037-1. URL https://www.sciencedirect.com/science/article/pii/S0893608005800371.

Matthew Trager, Kathlén Kohn, and Joan Bruna. Pure and spurious critical points: a geometric study of linear networks. In *ICLR*, 2020. URL https://openreview.net/forum?id=rkgOlCVYvB.

Konstantin Usevich, Ricardo Augusto Borsoi, Clara Dérand, and Marianne Clausel. Identifiability of deep polynomial neural networks. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=MrUsZfQ9pC.

Mohsen Yavartanoo, Shih-Hsuan Hung, Reyhaneh Neshatavar, Yue Zhang, and Kyoung Mu Lee. Polynet: Polynomial neural network for 3d shape recognition with polyshape representation. In *2021 International Conference on 3D Vision (3DV)*, pp. 1014–1023, 2021. doi: 10.1109/3DV53792.2021.00109.

AI DISCLOSURE SECTION

In this work, we used generative AI tools for editing texts and proofs aiming for clarity and conciseness. We also used generative AI tools in developing the package used in our numerical verification. LLM-generated code was verified by 1 author. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

A ROAD MAP TO THE APPENDICES

We provide a short description of what each appendix contain and what it contributes to the paper.

- Appendix A: Examples of convolution and Kronecker products of filters on small groups. The main purpose of this appendix is to develop the reader's intuition for group filters.
- Appendix B: Complete proofs for Propositions 2.6 and 2.7, the propositions that allow us to relate the parametrization maps φ and Φ.
- Appendix C: All main results in this paper are proved for general filters. In this appendix we define general filter, formally. In addition, we define symbolic ranks and explain their relation to general conditions. These concepts are necessary for the technical proofs in the later appendices.
- Appendix D: We prove that a general filter is invertible in the graded group algebra. This result is used repeatedly in the technical proofs to invert filters.
- Appendix E: A proof of Theorem 4.1 restricted to the rational map $\tilde{\varphi}$.
- Appendix F: A proof of Theorem 4.1 restricted to the rational map $\tilde{\Phi}$.
- Appendix G: A proof of Proposition 4.4.
- Appendix H: Proofs of Theorems 4.3 and 4.5.
- Appendix I: Numerical verification of our results for special cases of PGCNNs.

A EXAMPLES

In Definition 2.1 of graded group algebras a filter $\theta \in S[G]$ is described as a function from G to S. Another useful way to describe this filter is to identify it with the formal linear combination

$$\theta \equiv \sum_{g \in G} \theta(g) g.$$

Then the convolution operation becomes just the product of such formal linear combinations. Note that if $G = (\mathbb{Z}, +)$, then the description of a finite filter as a formal linear combination identifies it with a polynomial in one variable, which is the identification of convolutional filters in linear CNN presented in (Kohn et al., 2022).

Example A.1. Let $G = \mathbb{Z}_3 = \{0, 1, 2\}$ be the cyclic group of order 3 (with addition modulo 3), and let $S = \mathbb{R}$ concentrated in degree 0. Then

$$\mathbb{R}[G] = \{\theta : \mathbb{Z}_3 \to \mathbb{R}\} \cong \mathbb{R}^3.$$

A filter $\theta \in \mathbb{R}[G]$ is determined by its values

$$\theta(0) = a, \qquad \theta(1) = b, \qquad \theta(2) = c,$$

and can be written as the formal linear combination

$$\theta = a\,0 + b\,1 + c\,2.$$

Similarly, let $\psi \in \mathbb{R}[G]$ be given by

$$\psi(0) = x, \qquad \psi(1) = y, \qquad \psi(2) = z, \quad \text{so that} \quad \psi = x\,0 + y\,1 + z\,2.$$

The convolution $\theta * \psi$ is the function $\mathbb{Z}_3 \to \mathbb{R}$ defined by

$$(\theta * \psi)(g) = \sum_{h \in \mathbb{Z}_3} \theta(g - h)\psi(h),$$

where subtraction is taken modulo 3, or it is the formal sum

$$\begin{aligned} \theta * \psi &= (a\,0 + b\,1 + c\,2)(x\,0 + y\,1 + z\,2) \\ &= (ax + bz + cy)0 + (ay + bx + cz)1 + (az + by + cx)2. \end{aligned}$$

We finish this appendix by giving a concrete example of a Kronecker product of filters.

Example A.2. Let $G = H = \mathbb{Z}_2 = \{0, 1\}$ and $S = \mathbb{R}$. Fix the ordered bases $\{0, 1\}$ of $\mathbb{R}[G]$ and $\{(0,0), (0,1), (1,0), (1,1)\}$ of $\mathbb{R}[G \times H]$.

Let $\theta \in \mathbb{R}[G]$ and $\psi \in \mathbb{R}[H]$ be given by

$$\theta(0) = a,\ \theta(1) = b, \qquad \psi(0) = x,\ \psi(1) = y.$$

A direct computation shows that left convolution by θ and ψ is represented by the matrices

$$\mathrm{Mat}_\theta = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \qquad \mathrm{Mat}_\psi = \begin{pmatrix} x & y \\ y & x \end{pmatrix}.$$

The Kronecker product of these matrices is

$$\mathrm{Mat}_\theta \otimes \mathrm{Mat}_\psi = \begin{pmatrix} ax & ay & bx & by \\ ay & ax & by & bx \\ bx & by & ax & ay \\ by & bx & ay & ax \end{pmatrix}.$$

On the other hand, the filter $\theta \otimes \psi \in \mathbb{R}[G \times H]$ satisfies $(\theta \otimes \psi)(g, h) = \theta(g)\psi(h)$, and convolution by $\theta \otimes \psi$ on $\mathbb{R}[G \times H]$ is represented, in the chosen basis, by the same matrix.

B Proofs of Propositions 2.6 and 2.7

Proposition (2.6)**.** *Let $\theta_1, \theta_2 \in S[G]$ and $\psi_1, \psi_2 \in S[H]$. Then $(\theta_1 \otimes \psi_1) * (\theta_2 \otimes \psi_2) = (\theta_1 * \theta_2) \otimes (\psi_1 * \psi_2)$. By induction it holds that for $\theta_i, \psi_i \in S[G_i]$ for $1 \le i \le r$*

$$(\theta_1 \otimes \ldots \otimes \theta_r) * (\psi_1 \otimes \ldots \otimes \psi_r) = (\theta_1 * \psi_1) \otimes \ldots \otimes (\theta_r * \psi_r).$$

Proof. Use definition of convolution, we get

$$\begin{aligned}
(\theta_1 \otimes \psi_1) * (\theta_2 \otimes \psi_2)(g, h) &= \sum_{g', h'} (\theta_1 \otimes \psi_1)(gg'^{-1}, hh'^{-1})(\theta_2 \otimes \psi_2)(g', h') \\
\text{(use definition of } \otimes) &= \sum_{g', h'} \theta_1(gg'^{-1})\psi_1(hh'^{-1})\theta_2(g')\psi_2(h') \\
\text{(reorder the terms)} &= \sum_{g', h'} \theta_1(gg'^{-1})\theta_2(g')\psi_1(hh'^{-1})\psi_2(h') \\
\text{(split the sum into product of two sums)} &= \left(\sum_{g'} \theta_1(gg'^{-1})\theta_2(g')\right)\left(\sum_{h'} \psi_1(hh'^{-1})\psi_2(h')\right) \\
&= (\theta_1 * \theta_2)(g)(\psi_1 * \psi_2)(h) = (\theta_1 * \theta_2) \otimes (\psi_1 * \psi_2)(g, h)
\end{aligned}$$

□

Proposition (2.7)**.** *Let $\theta_1, \psi_1, \theta_2, \psi_2 \in S[G]$. Then*

$$(\theta_1 * \psi_1) \odot (\theta_2 * \psi_2) = ((\theta_1 \otimes \theta_2) * (\psi_1 \otimes \psi_2))\,|_G.$$

It follows that for $\theta, \psi \in S[G]$ and $r \ge 1$ we get $\sigma_r(\theta * \psi) = (\theta^{\otimes r} * \psi^{\otimes r})|_G$, where G is identified with the diagonal subgroup in G^r.

Proof. By Proposition 2.6,

$$(\theta_1 \otimes \theta_2) * (\psi_1 \otimes \psi_2) = (\theta_1 * \psi_1) \otimes (\theta_2 * \psi_2).$$

Restricting the right hand side to the diagonal subgroup $G < G^2$ gives us $(\theta_1 * \psi_1) \odot (\theta_2 * \psi_2)$, as required. □

C General conditions and symbolic ranks

To compute the dimension of the neuromanifold, which is the image of the parametrization map Φ (see §3), we will need to compute the rank of the Jacobian of the parametrization map Φ at a general set of parameters θ. In this appendix, we define what a general condition is, and we introduce the concept of symbolic matrix and symbolic ranks that will help us specify the general conditions necessary for our proofs.

Definition C.1 (generality)**.** We say that a condition $\mathcal{P}$ is *general* in $\mathbb{K}^m$ if it holds over a dense set in $\mathbb{K}^m$ equipped with the Zariski topology.

Another more intuitive way to understand generality, although not exactly equivalent, is to think of $\mathcal{P}$ as holding everywhere in $\mathbb{K}^m$ except for a set of Lebesgue measure zero.

Definition C.2. A symbolic matrix M is a matrix over a ring of polynomials $\mathbb{K}[y_1, \dots, y_m]$. We say M is symbolic in terms of the indeterminates $(y_1, \dots, y_m)$. The symbolic rank of M is decided by the vanishing of its minors in the ring $\mathbb{K}[y_1, \dots, y_m]$.

For a symbolic matrix M over $\mathbb{K}[y_1, \dots, y_m]$ and a point $p = (p_1, \dots, p_m) \in \mathbb{K}^m$, we construct the matrix $M(p)$ from M by substituting every y_i for p_i. Then the condition on p to satisfy $\text{rank}(M) = \text{rank}(M(p))$ is general in $\mathbb{K}^m$, because this condition fails if and only if p is the zero of all minors of M of size $\text{rank}(M) \times \text{rank}(M)$. We illustrate this in the following example.

Example C.3. Consider the symbolic matrix over $\mathbb{K}[y_1, y_2]$

$$M = \begin{pmatrix} y_1 & y_2 \\ y_2 & y_1 \end{pmatrix}.$$

The determinant of M is the polynomial

$$\det M = y_1^2 - y_2^2.$$

Hence, over the polynomial ring $\mathbb{K}[y_1, y_2]$, the symbolic rank of M is 2 since the determinant is not the zero polynomial in $\mathbb{K}[y_1, y_2]$. Therefore, a point $p = (p_1, p_2) \in \mathbb{K}^2$ satisfies $\text{rank}(M) = \text{rank}(M(p_1, p_2))$ if and only if $p_1^2 - p_2^2 \neq 0$ which is a general condition in $\mathbb{K}^2$. This example illustrates that the rank of a symbolic matrix is determined by polynomial conditions on the indeterminates and that rank deficiencies occur precisely on algebraic varieties defined by the vanishing of suitable minors.

D Invertibility of general filters

In this appendix, a general filter $\theta \in S[G]$ is shown to be invertible in the group algebra. Similar results hold for its extension to G^r, its rth Hadamard product, and its rth Kronecker product. These results will be relevant in the proof of the main theorems. In particular, since there we work in a general setting, we will be allowed to assume that the filters we work with are invertible in the group algebra.

Our first step toward this result is to establish the equivalence between invertibility of a filter $\theta \in S[G]$ with the invertability of its associated matrix Mat_θ. Notice that one direction of the equivalence is trivial. If $\theta \in S[G]$ is invertible then $\text{Mat}_\theta \text{Mat}_{\theta^{-1}} = I$. However, the other direction needs a little more work.

Remark D.1. Since Mat is the algebra representation ρ under a choice of base, it follows immediately that for any filters $\theta, \psi \in S[G]$ we get

$$\text{Mat}_\theta \text{Mat}_\psi = \text{Mat}_{\theta * \psi} \;\text{ and }\; \text{Mat}_\theta + \text{Mat}_\psi = \text{Mat}_{\theta + \psi}.$$

Also, the set $\{\text{Mat}_g \mid g \in G\}$ defines a basis for G-circulant matrices over the algebra S, where we identify $g \in G$ with the filter $g(h) = 1$ if $h = g$ otherwise $g(h) = 0$.

Proposition D.2. *Let $\theta \in S[G]$. If Mat_θ is invertible, then its inverse is also G-circulent. By Remark D.1, Mat_θ is invertible if and only if θ is invertible in $S[G]$.*

Proof. Since S is commutative, we can apply Cayley-Hamilton to the inverse Mat_θ^{-1} to write it as a polynomial in Mat_θ. Since $\text{Mat}_\theta = \sum_{g\in G} \theta(g)\text{Mat}_g$, then Mat_θ^{-1} is also a linear combination of the matrices Mat_g's with coefficients in S. Therefore Mat_θ^{-1} is also G-circulent. □

The second step toward our goal is to show that a general G-circulent matrix Mat_θ is invertible. To comment on what is aimed for in these proofs, note that a general square matrix M has, of course, a full rank because its determinant defines a nontrivial polynomial in terms of its entries. What we show in what follows is that for G-circulent matrices, which have weight-sharing constraints, the determinant still defines a nontrivial polynomial regardless of the group G.

Definition D.3. For a filter $\theta \in S[G]$, we define its determinant to be the determinant of its associated matrix, i.e. $\det(\theta) := \det(\text{Mat}_\theta)$.

In the following proposition, we relate the determinants of the three filters $\theta \in S[G]$ and $\theta^{\otimes}, \theta^{G^r} \in S[G^r]$.

Proposition D.4. *Let $\theta \in S[G]$, then we have*

$$\det(\theta^{\otimes r}) = \det(\theta)^{rn^{r-1}} \quad \textit{and} \quad \det(\theta^{G^r}) = \det(\theta)^{n^{r-1}}$$

where $n = |G|$. This implies that $\theta \in S[G]$ is invertible if and only if $\theta^{\otimes r}$ and θ^{G^r} in $S[G^r]$ are both invertible.

Proof. The first equality follows directly from the property of the Kronecker product for matrices.

The second equality follows from a standard argument in representation theory. Since we identify G with the diagonal subgroup $\Delta < G^r$, we obtain that the restriction of the regular algebra representation of $S[G^r]$ to $S[G]$ has $|G^r/G| = n^{r-1}$ unique $S[G]$-invariant sub-representation $V_i \subset S[G^r]$ where $S[G^r] = \bigoplus_{i=1}^{n^{r-1}} V_i$ and $S[G] \cong V_i$ for all i. Therefore, under a suitable choice of basis, we have $\text{Mat}_\theta^{G^r}$ is a block diagonal matrix of n^{r-1} blocks where each block is equal to Mat_θ. This proves our claim. □

In the following proposition, we state that invertibility of filters is a general condition in $S[G]$.

Proposition D.5. *Let $\theta \in S[G]$ then the following conditions are general for all $r \geq 1$:*

1. $\det(\theta) \neq 0$ *2.* $\det(\theta^{G^r}) \neq 0$ *3.* $\det(\theta^{\otimes r}) \neq 0$ *4.* $\det(\sigma_r(\theta)) \neq 0$.

Proof. Assume $G = \{g_1, \ldots, g_n\}$ and $\theta = (\theta_{g_1}, \ldots, \theta_{g_n})$ where $g_1 = e$ is the group identity. We have $(\text{Mat}_\theta)_{i,j} = \theta(g_i g_j^{-1})$. For fixed g_i (row) and g_k there is a unique g_j such that $g_i g_j^{-1} = g_k$, namely $g_j = g_k^{-1} g_i$. Similarly for fixed g_j (column) and g_k there is a unique g_i such that $g_i g_j^{-i} = g_k$. This means the element $\theta(g_k)$ occurs once and only once in each column and in each row in the matrix Mat_θ. Therefore by elementary rules of the determinant, the term $\theta(g_k)^n$ for all $g_k \in G$ occurs once and only once in the determinant of Mat_θ. This implies the condition $\det(\text{Mat}_\theta) = \det(\theta) \neq 0$ in $S[G]$ is general. By Proposition D.4, the conditions 2 and 3 are general.

The same argument shows that $\det(\sigma_r(\theta))$ is not the trivial polynomial because it includes the terms $\theta(g_k)^{rn}$ for all $g_k \in G$. □

Corollary D.6. *For $S = \mathbb{K}$, Propositions D.4 and D.5 imply that generally $\theta, \sigma_r(\theta) \in \mathbb{K}[G]$ and $\theta^{\otimes r}, \theta^{G^r} \in \mathbb{K}[G^r]$ are invertible, for all $r \geq 1$. In particular, they are generally not zero divisors in their algebras.*

E REGULARITY OF $\tilde{\varphi}$

In this appendix we provide a complete proof of the following theorem which is Theorem 4.1 restricted to the map $\tilde{\varphi}$.

Theorem E.1. *For activation degree $r \geq 2$, the rational map $\tilde{\varphi}$ is* generally regular, *i.e. its Jacobian is full rank at a general point in its domain. This implies that* $\dim(\mathcal{M}_\varphi) = L(n-1)+1$*, where $n = |G|$ and L is the number of layers.*

Proof. For $\theta := (\theta_1, \ldots, \theta_n)$ be general in $\mathbb{K}^{nL}$. We have $T_{[\theta]}(\mathbb{P}(\mathbb{K}^n))^L = \bigoplus_i \mathbb{K}^n/\mathbb{K}\theta_i$, and if $[\dot{\theta}] \in \ker J_{[\theta]}\tilde{\varphi}$ then $J_\theta\varphi(\dot{\theta}) \in \mathbb{K}\varphi(\theta)$. By Euler theorem for homogeneous functions, we get that $J_\theta\varphi(\theta_1, 0, \ldots, 0) = r^{L-1}\varphi(\theta)$, and therefore $\dot{\theta} + k(\theta_1, 0, \ldots, 0) \in \ker J_\theta\varphi$, for some $k \in \mathbb{K}$, which by Proposition E.2 is contained in $\bigoplus_i \mathbb{K}\theta_i$. Hence $[\dot{\theta}] = 0 \in T_{[\theta]}(\mathbb{P}(\mathbb{K}^n))^L$ and $J_{[\theta]}\tilde{\varphi}$ is full rank. □

The proof of Theorem E.1 relies strongly on the next proposition.

Proposition E.2. *Let L be the number of layers and r be the degree of activation. Let also $\theta = (\theta_1, \ldots, \theta_L)$ denote the tuple of filters that define φ_θ. Then for a general θ, the kernel of the Jacobian $J_\theta\varphi$ has a basis of $L-1$ elements given by*

$$((\theta_1, -r\theta_2, 0, \ldots, 0), (0, \theta_2, -r\theta_3, 0, \ldots, 0), \ldots, (0, \ldots, 0, \theta_{L-1}, -r\theta_L)).$$

Before we state the proof of Proposition E.2, we need to set some notation that applies only in this appendix. We also need to prove two lemmata, one easy and one hard.

Notation.

- We introduce the following notation

$$[\breve{x}_1 \diamond x_2 \diamond \ldots \diamond x_n] := \breve{x}_1 \diamond x_2 \diamond \ldots \diamond x_n \; + \; x_1 \diamond \breve{x}_2 \diamond \ldots \diamond x_n \; + \ldots + \; x_1 \diamond \ldots \diamond \breve{x}_n,$$

  where $\breve{}$ and $\diamond$ denote arbitrary operations.
- Let $g \in G^r$ and $h \in G$ we define $g \cdot h := g \cdot h^r$.

Lemma E.3. *Let θ and ψ be G-filters, then*

1. $$J_{\theta,\psi}(\theta * \psi)(\dot{\theta}, \dot{\psi}) = \dot{\theta} * \psi + \theta * \dot{\psi} = [\dot{\theta} * \psi].$$

2. $$\begin{aligned} J_\theta(\theta^{\otimes r})(\dot{\theta}) &= \dot{\theta} \otimes \theta \otimes \ldots \otimes \theta + \theta \otimes \dot{\theta} \otimes \ldots \otimes \theta + \ldots + \theta \otimes \ldots \otimes \theta \otimes \dot{\theta} \\ &= \dot{\theta} \otimes \theta^{\otimes(r-1)} + \theta \otimes \dot{\theta} \otimes \theta^{\otimes(r-2)} + \ldots + \theta^{\otimes(r-1)} \otimes \dot{\theta} \\ &= [\dot{\theta} \otimes \theta \otimes \ldots \otimes \theta]. \end{aligned}$$

3. *Let $\theta = (\theta_1, \ldots, \theta_L)$, $\dot{\theta} = (\dot{\theta}_1, \ldots, \dot{\theta}_L)$. Moreover, let θ' and $\dot{\theta}'$ be the tuples containing the first $L-1$ filters in θ and $\dot{\theta}$, respectively. Since $\varphi_\theta = \varphi_{\theta'}^{\otimes r} * \theta_L|^{G^{rL-1}}$, we get*

$$J_\theta\varphi(\dot{\theta}) = [J_{\theta'}\varphi(\dot{\theta}') \otimes \underbrace{\varphi_{\theta'} \otimes \ldots \otimes \varphi_{\theta'}}_{=\varphi_{\theta'}^{\otimes(r-1)}}] * \theta_L + \varphi_{\theta'}^{\otimes r} * \dot{\theta}_L.$$

Proof. 1.

$$J_{\theta,\psi}(\theta * \psi)(\dot{\theta}, \dot{\psi}) = J_{\theta,\psi}\left(\left(\sum_h \theta(g_1 \cdot h^{-1})\psi(h), \ldots, \sum_h \theta(g_n \cdot h^{-1})\psi(h)\right)\right)(\dot{\theta}, \dot{\psi}).$$

Since $J_{\theta,\psi}\left(\sum_h \theta(g_i \cdot h^{-1})\psi(h)\right)(\dot{\theta}, \dot{\psi}) = \sum_h \dot{\theta}(g_i \cdot h^{-1})\psi(h) + \sum_h \theta(g_i \cdot h^{-1})\dot{\psi}(h)$, we get

$$J_{\theta,\psi}(\theta * \psi)(\dot{\theta}, \dot{\psi}) = \dot{\theta} * \psi + \theta * \dot{\psi}.$$

2. We prove this for $r = 2$, the general case follows by induction.

   We have $\theta \otimes \theta = (\theta_1^2, \theta_1\theta_2, \ldots, \theta_1\theta_n, \theta_2\theta_1, \ldots, \theta_2\theta_n, \ldots, \theta_n\theta_1, \ldots, \theta_n^2)$, hence $J_\theta(\theta \otimes \theta)$ is a $n^2 \times n$ matrix where the $((i-1)n+j)$th row corresponding to the function $\theta_i\theta_j$ will have θ_j at the ith column, θ_i a the jth column, and zero everywhere else. Therefore, the $((i-1)n+j)$th row in $J_\theta(\theta \otimes \theta)(\dot{\theta})$ is equal to $\theta_i\dot{\theta}_j + \dot{\theta}_i\theta_j$.

3. Direct application of the first rule, then the second rule.

□

Lemma E.4. *Let $\theta = (\theta_1, \ldots, \theta_{L-1}, e) \in \mathbb{K}[G]^L$ where θ_i is a general filter for $1 \leq i < L$ and $e \in G \subset \mathbb{K}[G]$ is the identity element. If $\dot{\theta} = (\dot{\theta}_1, \ldots, \dot{\theta}_L) \in \ker J_\theta\varphi$ then $\dot{\theta}_L = \lambda e$ for some $\lambda \in \mathbb{K}$.*

Proof. Write $\theta' = (\theta_1, \ldots, \theta_{L-1})$ and similarly $\dot{\theta}' = (\dot{\theta}_1, \ldots, \dot{\theta}_{L-1})$. For brevity, write,

$$J := J_{\theta'}\varphi(\dot{\theta}').$$

By Lemma E.3 (the Jacobian computation rules), $\dot{\theta} \in \ker\big(J_\theta\varphi\big)$ if and only if

$$\big[J \otimes \varphi_{\theta'}^{\otimes(r-1)}\big] \;=\; -\,\varphi_{\theta'}^{\otimes r} * \dot{\theta}_L \tag{3}$$

Evaluate (3) at tuples of the special form $\tilde{g}_i = (g_i, \ldots, g_i) \in G^{r^{L-1}}$ (where $g_i \in G^{r^{L-2}}$). This yields, for each i,

$$r\,(\varphi_{\theta'}(g_i))^{r-1} J(g_i) \;=\; \sum_{h \in G} (\varphi_{\theta'}(g_i h^{-1}))^r \dot{\theta}_L(h).$$

Hence

$$J(g_i) \;=\; \frac{1}{r}\,\varphi_{\theta'}(g_i)\,\dot{\theta}_L(e) \;+\; \frac{1}{r} \sum_{h \in G\setminus\{e\}} \frac{(\varphi_{\theta'}(g_i h^{-1}))^r}{(\varphi_{\theta'}(g_i))^{\,r-1}}\,\dot{\theta}_L(h). \tag{4}$$

Now evaluate (3) at $\tilde{g} = (g_{i_1}, \ldots, g_{i_r})$ to get,

$$\sum_{k=1}^{r} J(g_{i_k}) \prod_{j \neq k} \varphi_{\theta'}(g_{i_j}) = \sum_{h \in G} \varphi_{\theta'}^{\otimes r}(\tilde{g}h^{-1})\dot{\theta}_L(h).$$

Substituting (4) in the above equation gives a linear system in the unknowns $\{\dot{\theta}_L(h)\}_{h \neq e}$ with coefficient matrix C whose entries are symbolic expressions in the values $\varphi_{\theta'}(g_i)$. Denote these coefficients by

$$c_{\tilde{g},h} = \frac{1}{r} \sum_{k=1}^{r} \left(\frac{(\varphi_{\theta'}(g_{i_k}h^{-1}))^r \prod_{j \neq k} \varphi_{\theta'}(g_{i_j})}{(\varphi_{\theta'}(g_{i_k}))^{r-1}} \right) - \varphi_{\theta'}^{\otimes r}(\tilde{g}h^{-1}).$$

Observe $c_{\tilde{g},e} = 0$ for every $\tilde{g}$ that is why we exclude the variable $\dot{\theta}_L(e)$ from the unknowns. The matrix C has $|G^{r^{L-1}}| = n^{r^{L-1}}$ rows and $n-1$ columns, so C is full column rank if and only if some $(n-1) \times (n-1)$ minor is nonzero as a symbolic polynomial in the variables θ'.

Choose the $(n-1)$ rows corresponding to

$$\tilde{g}^{\,i} := (g_1, g_i, \ldots, g_i) \qquad (2 \leq i \leq n),$$

where $g_j = (h_j, \ldots, h_j) \in G^{r^{L-2}}$, $h_1 = e$ and $h_j \in G$. For this choice, one checks that the diagonal entries of the resulting $(n-1)$-minor contain a factor $(\varphi_{\theta'}(g_1))^{r+1}$ which does not appear elsewhere in that minor. Namely,

$$c_{\tilde{g}^i,h_j} = \frac{1}{r}\frac{(\varphi_{\theta'}(g_i))^{r-1}}{(\varphi_{\theta'}(g_1))^{r-1}}\varphi_{\theta'}((g_1h_j^{-1}))^r + \frac{r-1}{r}\frac{(\varphi_{\theta'}(g_ih^{-j}))^r}{\varphi_{\theta'}(g_i)}\varphi_{\theta'}(g_1) - \varphi_{\theta'}^{\otimes r}(\tilde{g}^i h_j^{-1}).$$

and

$$c_{\tilde{g}^i,h_i} = \frac{1}{r}\frac{(\varphi_{\theta'}(g_i))^{r-1}}{(\varphi_{\theta'}(g_1))^{r-1}}\varphi_{\theta'}((g_1h_i^{-1}))^r + \frac{r-1}{r}\frac{(\varphi_{\theta'}(g_1))^{r+1}}{\varphi_{\theta'}(g_i)} - \varphi_{\theta'}^{\otimes r}(\tilde{g}^i h_i^{-1}).$$

Since $\varphi_{\theta'}(g_1)$ has monomials in the parameters of θ' that do not occur in $\varphi_{\theta'}(g_i)$ for $2 \geq i \leq n$, then this symbolic minor is nonvanishing for general θ', so C is symbolically full rank. Therefore the only solution is $\dot{\theta}_L(h) = 0$ for all $h \neq e$. □

We are finally ready to give a proof of Proposition E.2.

Proof of Proposition E.2. Let the perturbations be $\dot{\theta} = (\dot{\theta}_1, \ldots, \dot{\theta}_L)$ and write $\theta' = (\theta_1, \ldots, \theta_{L-1})$ and similarly $\dot{\theta}' = (\dot{\theta}_1, \ldots, \dot{\theta}_{L-1})$. We prove the claim by induction on L.

Base case $L = 1$**.** If $L = 1$ then $\varphi_\theta = \theta_1$ and

$$J_\theta \varphi(\dot{\theta}) = J_{\theta_1}(\theta_1)(\dot{\theta}_1) = \dot{\theta}_1,$$

so $J_\theta \varphi$ is the identity and hence full rank. The kernel is trivial, which matches the statement for $L = 1$.

Inductive step. Assume the proposition holds for $L-1$ layers; we prove it for L layers. For brevity, write,

$$J := J_{\theta'} \varphi(\dot{\theta}').$$

By Lemma E.3 (the Jacobian composition / convolution rule), $\dot{\theta} \in \ker\big(J_\theta \varphi\big)$ if and only if

$$\big[J \otimes \varphi_{\theta'}^{\otimes(r-1)}\big] * \theta_L \;=\; -\varphi_{\theta'}^{\otimes r} * \dot{\theta}_L. \tag{5}$$

If $\dot{\theta}_L = 0$ then $J = 0$ the result follows by the inductive step. Assume $\dot{\theta}_L \neq 0$, convolute both sides of (5) with the inverse of θ_L, which exists because θ_L is general, and let $\dot{\tilde{\theta}}_L := \dot{\theta}_L * \theta_L^{-1}$ to get

$$\big[J \otimes \varphi_{\theta'}^{\otimes(r-1)}\big] \;=\; -\varphi_{\theta'}^{\otimes r} * \dot{\tilde{\theta}}_L. \tag{6}$$

Since θ' is general and $\dot{\theta}_L \neq 0$, then by Lemma E.4, we have $\dot{\tilde{\theta}}_L = \lambda e$ for some nonzero $\lambda \in \mathbb{K}$. This implies $\dot{\theta}_L = \lambda \theta_L$, and Therefore, any $\dot{\theta} \in \ker J_\theta \varphi$ must have the form

$$\dot{\theta} = (\dot{\theta}_1, \ldots, \dot{\theta}_{L-1},\ \lambda \theta_L).$$

Observe that the vector

$$(0, \ldots, 0,\ -\tfrac{\lambda}{r}\theta_{L-1},\ \lambda\theta_L)$$

lies in the kernel as well (it is a direct check using the Jacobian composition rule). Subtracting this vector from $\dot{\theta}$ yields

$$(\dot{\theta}_1, \ldots, \dot{\theta}_{L-1} - (-\tfrac{\lambda}{r}\theta_{L-1}),\ 0) \;=\; (\dot{\theta}_1, \ldots, \dot{\theta}_{L-1} + \tfrac{\lambda}{r}\theta_{L-1},\ 0) \in \ker J_\theta \varphi.$$

By Eq. (5), the vector

$$(\dot{\theta}_1, \ldots, \dot{\theta}_{L-1} + \tfrac{\lambda}{r}\theta_{L-1}) \in \ker J_{\theta'} \varphi,$$

and by the inductive hypothesis, it can be written as a linear combination of

$$\big((\theta_1, -r\theta_2, 0, \ldots, 0),\ (0, \theta_2, -r\theta_3, 0, \ldots, 0),\ \ldots,\ (0, \ldots, 0, \theta_{L-2}, -r\theta_{L-1})\big).$$

Adding back the previously subtracted vector $(0, \ldots, 0, -\frac{\lambda}{r}\theta_{L-1}, \lambda\theta_L)$ shows that $\dot{\theta}$ is a linear combination of the $L-1$ vectors stated in the proposition. This completes the induction and the proof. □

F Regularity of $\tilde{\Phi}$

In this appendix we provide a complete proof of the following theorem which is Theorem 4.1 restricted to the map $\tilde{\Phi}$.

Theorem F.1. *For activation degree $r \geq 2$, the rational map $\tilde{\Phi}$ is* generally regular*; that is, its Jacobian has full rank at a general point of its domain. This implies* $\dim(\mathcal{M}_\Phi) = L(n-1) + 1$.

Proof. The claim follows directly from Proposition F.2, in the same manner that Proposition E.2 implies the general regularity of $\tilde{\varphi}$ in Theorem E.1. □

Proposition F.2. *Let L be the number of layers and r the activation degree. Let $\theta = (\theta_1, \dots, \theta_L)$ denote the tuple of filters defining Φ_θ. Then for a general θ, the kernel of the Jacobian $J_\theta\Phi$ has an $(L-1)$–dimensional basis given by*

$$\big((\theta_1, -r\theta_2, 0, \dots, 0),\ (0, \theta_2, -r\theta_3, 0, \dots, 0),\ \dots,\ (0, \dots, 0, \theta_{L-1}, -r\theta_L) \big).$$

Before we give a complete proof of Proposition F.2, we will need to introduce new notation and two lemmata.

Notation. Recall that a polynomial G-convolutional neural network of depth L, and activation r is the map

$$\begin{aligned} \Phi : \mathbb{K}[G]^L &\to \mathrm{Sym}_{\mathbb{K}}(n, r^{L-1})[G] \\ (\theta_1, \dots, \theta_L) &\mapsto \sigma_r\Big(\cdots \sigma_r\big(\sigma_r(x * \theta_1) * \theta_2\big) \cdots * \theta_{L-1} \Big) * \theta_L. \end{aligned}$$

Equivalently, the map Φ can be defined recursively by the formula

$$\Phi(\theta_1, \dots, \theta_L) = \sigma_r\big(\Phi(\theta_1, \dots, \theta_{L-1})\big) * \theta_L.$$

Recall also from Section 3 that the map Φ can also be defined in terms of the map φ by the formula

$$\Phi(\theta_1, \dots, \theta_L) = \big(x^{r^{L-1}} * \varphi(\theta_1, \dots, \theta_L)\big)\big|_G. \tag{7}$$

For $\theta := (\theta_1, \dots, \theta_L) \in \mathbb{K}[G]^L$, define the filter Φ_θ to be the one obtained by evaluating Φ at θ i.e.

$$\begin{aligned} \Phi_\theta : G &\to \mathrm{Sym}_{\mathbb{K}}(n, r^{L-1}) \\ g &\mapsto \Phi(\theta)(g). \end{aligned}$$

Each monomial in $\Phi_\theta(g) \in \mathrm{Sym}_{\mathbb{K}}(n, r^{L-1})$ is indexed by an unordered tuple

$$[g_1, \dots, g_{r^{L-1}}] \in G^{r^{L-1}},$$

or, equivalently, by multiplicities

$$[r_1 g_1, \dots, r_j g_j], \qquad r_1 + \cdots + r_j = r^{L-1}.$$

We denote the coefficient of this monomial in $\Phi_\theta(g)$ by

$$\Phi_\theta(g)[r_1 g_1, \dots, r_j g_j],$$

or simply $\Phi_\theta[r_1 g_1, \dots, r_j g_j]$ when the group element g is clear from context.

Lemma F.3. *Let $\theta = (\theta_1, \dots, \theta_L)$ and $\dot{\theta} = (\dot{\theta}_1, \dots, \dot{\theta}_L)$, and set*

$$\theta' = (\theta_1, \dots, \theta_{L-1}), \qquad \dot{\theta}' = (\dot{\theta}_1, \dots, \dot{\theta}_{L-1}).$$

Then:

1. *For any index $[g_1, \dots, g_{r^{L-1}}]$ we have*

$$\Phi_\theta(g)[g_1, \dots, g_{r^{L-1}}] = \sum_{\delta \in S_{r^{L-1}}} \varphi_\theta(g_{\delta(1)}^{-1} g, \dots, g_{\delta(r^{L-1})}^{-1} g) = \Phi_\theta(e)[g^{-1} g_1, \dots, g^{-1} g_{r^{L-1}}], \tag{8}$$

   where S_m denotes the permutation group on m elements.

2. *As proved in Shahverdi et al. (2025),*

$$J_\theta \Phi(\dot{\theta}) = r\big(J_{\theta'}\Phi(\dot{\theta}') \odot \sigma_{r-1}(\Phi_{\theta'})\big) * \theta_L \ + \ \sigma_r(\Phi_{\theta'}) * \dot{\theta}_L.$$

   The Jacobian of Φ_θ is obtained by computing the Jacobian of each polynomial coefficient of $\Phi_\theta(g)$ for all $g \in G$.

Proof. Using (7), for any $g \in G$,

$$\Phi_\theta(g) = \sum_{h \in G^{r^{L-1}}} x^{r^{L-1}}(h)\, \varphi_\theta(h^{-1}g).$$

Since $x^{r^{L-1}}(h_1) = x^{r^{L-1}}(h_2)$ whenever h_1, h_2 differ by a permutation, the formula (8) follows immediately. □

In other words, the first part of Lemma F.3 states that the coefficients of the polynomial $\Phi_\theta(g)$ are completely determined by the coefficients of the polynomial $\Phi_\theta(e)$ at the group identity $e \in G$.

Lemma F.4. *Let $\theta = (\theta_1, \dots, \theta_{L-1}, e) \in \mathbb{K}[G]^L$ where θ_i is a general filter for $1 \le i < L$ and $e \in G \subset \mathbb{K}[G]$ is the identity element. If $\dot\theta = (\dot\theta_1, \dots, \dot\theta_L) \in \ker J_\theta\Phi$ then $\dot\theta_L = \lambda e$ for some $\lambda \in \mathbb{K}$.*

Proof. Let $G = \{g_1 = e, g_2, \dots, g_n\}$, $\theta' = (\theta_1, \dots, \theta_{L-1})$ and $\dot\theta' = (\dot\theta_1, \dots, \dot\theta_{L-1})$. For brevity set

$$J := J_{\theta'}\Phi(\dot\theta').$$

Lemma F.3 implies that $\dot\theta \in \ker J_\theta\Phi$ if and only if

$$rJ \bigodot \sigma_{r-1}(\Phi_{\theta'}) = -\,\sigma_r(\Phi_{\theta'}) * \dot\theta_L. \tag{9}$$

Evaluate the filters on both sides in Eq. (9) at $g = e$ to get

$$(rJ \bigodot \sigma_{r-1}(\Phi_{\theta'}))(e) = -\sum_{h \in G} \sigma_r(\Phi_{\theta'})(h)\dot\theta_L(h^{-1}). \tag{10}$$

Equation (10) is an equality of two homogeneous polynomials in $\mathrm{Sym}_{\mathbb{K}}(x, r^{L-1})$. Therefore for an index $[r_1g_1, \dots, r_ng_n]$ with $r_1 + \cdots + r_n = r^{L-1}$, equation (10) gives the equation

$$(rJ \bigodot \sigma_{r-1}(\Phi_{\theta'}))(e)[r_1g_1, \dots, r_ng_n] = -\sum_{h \in G} \sigma_r(\Phi_{\theta'})(h)[r_1g_1, \dots, r_ng_n]\dot\theta_L(h^{-1}),$$

which by Proposition F.3 is equivalent to

$$(rJ \bigodot \sigma_{r-1}(\Phi_{\theta'}))(e)[r_1g_1, \dots, r_ng_n] = -\sum_{h \in G} \sigma_r(\Phi_{\theta'})(e)[r_1hg_1, \dots, r_nhg_n]\dot\theta_L(h).$$

Since the above equation is a relation between the polynomials at only e, we remove e from the notation and write

$$(rJ \bigodot \sigma_{r-1}(\Phi_{\theta'}))[r_1g_1, \dots, r_ng_n] = -\sum_{h \in G} \sigma_r(\Phi_{\theta'})[r_1hg_1, \dots, r_nhg_n]\dot\theta_L(h). \tag{$*$}$$

These equations $(*)$ form a system of equations in the unknowns $\dot\theta$; one equation for each multindex $[r_1g_1, \dots, r_ng_n]$. Note that on the one hand an unknown appears in the homogeneous polynomial J if and only if it comes from the first $L-1$ filters in $\dot\theta$. On the other hand, the claim of this lemma concerns only the unknowns $\dot\theta_L$. Therefore our goal is to eliminate the unknown homogeneous polynomial J from the system $(*)$. To achieve this, we need to express the coefficient of J in terms of $\dot\theta_L$. We do this step by step as follows.

First coefficients. For the index $[r^{L-1}g_i]$, Eq $(*)$ becomes

$$r(J \bigodot \sigma_{r-1}(\Phi_{\theta'}))[r^{L-1}g_i] = -\sum_{h \in G} \sigma_r(\Phi_{\theta'})[r^{L-1}hg_i]\,\dot\theta_L(h).$$

Unfolding the Hadamard product on the left hand side gives

$$r(J \bigodot \sigma_{r-1}(\Phi_{\theta'}))[r^{L-1}g_i] = rJ[r^{L-2}g_i]\,\sigma_{r-1}(\Phi_{\theta'})[(r-1)r^{L-2}g_i].$$

So we obtain

$$J[r^{L-2}g_i] = -\sum_{h \in G} \frac{\sigma_r(\Phi_{\theta'})[r^{L-1}hg_i]}{r\,\sigma_{r-1}(\Phi_{\theta'})[(r-1)r^{L-2}g_i]}\dot\theta_L(h). \tag{$\dagger$}$$

Next coefficients. For the multiindex $[g_1, (r^{L-1}-1)g_i]$,

$$r(J \odot \sigma_{r-1}(\Phi_{\theta'}))[g_1, (r^{L-1}-1)g_i] = -\sum_{h\in G} \sigma_r(\Phi_{\theta'})[hg_1, (r^{L-1}-1)hg_i]\,\dot\theta_L(h).$$

The left side expands as

$$rJ[g_1, (r^{L-2}-1)g_i]\,\sigma_{r-1}(\Phi_{\theta'})[(r-1)r^{L-2}g_i] + rJ[r^{L-2}g_i]\,\sigma_{r-1}(\Phi_{\theta'})[g_1, ((r-1)r^{L-2}-1)g_i],$$

so together with (†) this expresses $J[g_1, (r^{L-2}-1)g_i]$ linearly in $\dot\theta_L$.

Iterating. Define for $0 \le m \le r^{L-2}$

$$O^i_m = \sum_{h\in G} \frac{\sigma_r(\Phi_{\theta'})[mhg_1, (r^{L-1}-m)hg_i]}{r\,\sigma_{r-1}(\Phi_{\theta'})[(r-1)r^{L-2}g_i]}\dot\theta_L(h),$$

and

$$D^i_k = \sigma_{r-1}(\Phi_{\theta'})[kg_1, ((r-1)r^{L-2}-k)g_i].$$

Define O_m to be the square $n\times n$ matrix where its rows are given by O^i_m. Namely, the entry (i,h) is given by

$$\frac{\sigma_r(\Phi_{\theta'})[mhg_1, (r^{L-1}-m)hg_i]}{r\,\sigma_{r-1}(\Phi_{\theta'})[(r-1)r^{L-2}g_i]}.$$

For $0 \le m < k \le r^{L-2}$ set

$$C^i_{m,k} = \sigma_{r-1}(\Phi_{\theta'})[(k-m)g_1, ((r-1)r^{L-2}+m-k)g_i] = D^i_{k-m},$$

and

$$C^{m,k}_i = \sum_{\{\alpha_1=m<\cdots<\alpha_f=k\}} \prod_{j=1}^{f-1} C^i_{\alpha_j,\alpha_{j+1}}, \qquad C^{k,k}_i = 1.$$

Induction on r_1 yields, for $0 \le r_1 \le r^{L-2}-1$ and $1 < i \le n$:

$$J[r_1 g_1, (r^{L-2}-r_1)g_i] = \sum_{m=0}^{r_1} O^i_m\, C^{m,r_1}_i. \qquad (\clubsuit)$$

Constructing the linear system only in $\dot\theta_L$. Using ($\clubsuit$) and (†), for each $i \ne 1$ the coefficient equation ($*$) corresponding to $[r^{L-2}g_1, (r-1)r^{L-2}g_i]$ gives

$$r\sum_{r_1=0}^{r^{L-2}} J[r_1 g_1, (r^{L-2}-r_1)g_i]\, D^i_{r^{L-2}-r_1} + rO^i_{r^{L-2}} D^i_0 = 0.$$

This is linear in $\dot\theta_L$, since each $J[\cdot]$ and O^i_m is linear in $\dot\theta_L$.

These $n-1$ equations (indexed by $i \ne 1$) form a $(n-1)\times n$ symbolic matrix in the parameters θ'. The entry in row i and column $h = g_i^{-1}$ contains the term

$$\big(\Phi_{\theta'}[r^{L-2}g_1])\big)^{r+1}\big(\Phi_{\theta'}[r^{L-2}g_i]\big)^{r-1},$$

which is a polynomial in the parameters θ' (the term appears in the sum where $r_1 = 0$). Excluding the entry at the column $h = e$, no other entry in this row contains a term that has the factor $\big(\Phi_{\theta'}[r^{L-2}g_1])\big)^{r+1}$.

Hence the submatrix omitting the column $h = e$ has a symbolically nonvanishing determinant, so the kernel is one–dimensional and spanned by $(1, 0, \ldots, 0)$. Therefore $\dot\theta_L(h) = 0$ for $h \ne e$. This completes the proof of the lemma. □

We are finally ready to give the proof of Proposition F.2

Proof of Proposition F.2. We write perturbations as $\dot{\theta} = (\dot{\theta}_1, \dots, \dot{\theta}_L)$. Set $\theta' = (\theta_1, \dots, \theta_{L-1})$ and $\dot{\theta}' = (\dot{\theta}_1, \dots, \dot{\theta}_{L-1})$. We proceed by induction on L.

Base case $L = 1$**.** For $L = 1$ we have $\Phi_\theta = x * \theta_1$, hence

$$J_\theta \Phi(\dot{\theta}) = J_{\theta_1}(x * \theta_1)(\dot{\theta}_1) = x * \dot{\theta}_1.$$

This vanishes iff $\dot{\theta}_1 = 0$, so the kernel is trivial, consistent with the proposition.

Inductive step. Assume the proposition holds for all networks with $L - 1$ layers. We prove it for L layers.

For brevity set

$$J := J_{\theta'} \Phi(\dot{\theta}').$$

Lemma F.3 implies that

$$\dot{\theta} \in \ker J_\theta \Phi \quad \Longleftrightarrow \quad r\big(J \odot \sigma_{r-1}(\Phi_{\theta'})\big) * \theta_L = -\sigma_r(\Phi_{\theta'}) * \dot{\theta}_L. \tag{11}$$

Let $\dot{\theta}_L = 0$. Since θ_L is general, so by Corollary D.6 we get θ_L is not a zero divisor, and hence $J = 0$. The required result follows by the inductive hypothesis.

Assume $\dot{\theta}_L \neq 0$, convolute both sides of (11) with the inverse of θ_L, which exists because θ_L is general, and let $\dot{\tilde{\theta}}_L := \dot{\theta}_L * \theta_L^{-1}$ to get

$$rJ \odot \Phi_{\theta'}^{r-1} \;=\; -\Phi_{\theta'}^{r} * \dot{\tilde{\theta}}_L. \tag{12}$$

Since θ' is general and $\dot{\theta}_L \neq 0$, then by Lemma F.4, we have $\dot{\tilde{\theta}}_L = \lambda e$ for some nonzero $\lambda \in \mathbb{K}$. This implies $\dot{\theta}_L = \lambda \theta_L$, and therefore any $\dot{\theta} \in \ker J_\theta \Phi$ must have the form

$$\dot{\theta} = (\dot{\theta}_1, \dots, \dot{\theta}_{L-1},\ \lambda \theta_L).$$

Observe that the vector

$$(0, \dots, 0,\ -\tfrac{\lambda}{r}\theta_{L-1},\ \lambda \theta_L)$$

lies in the kernel as well (it is a direct check using the Jacobian composition rule). Subtracting this vector from $\dot{\theta}$ yields

$$(\dot{\theta}_1, \dots, \dot{\theta}_{L-1} - (-\tfrac{\lambda}{r}\theta_{L-1}),\ 0) \;=\; (\dot{\theta}_1, \dots, \dot{\theta}_{L-1} + \tfrac{\lambda}{r}\theta_{L-1},\ 0) \in \ker J_\theta \Phi.$$

By Eq. (11) the vector

$$(\dot{\theta}_1, \dots, \dot{\theta}_{L-1} + \tfrac{\lambda}{r}\theta_{L-1}) \in \ker J_{\theta'} \Phi,$$

and by the inductive hypothesis, it can be written as a linear combination of

$$\big((\theta_1, -r\theta_2, 0, \dots, 0),\ (0, \theta_2, -r\theta_3, 0, \dots, 0),\ \dots,\ (0, \dots, 0, \theta_{L-2}, -r\theta_{L-1})\big).$$

Adding back the previously subtracted vector $(0, \dots, 0, -\frac{\lambda}{r}\theta_{L-1}, \lambda\theta_L)$ shows that $\dot{\theta}$ is a linear combination of the $L - 1$ vectors stated in the proposition. This completes the induction and the proof.

□

G Proof of Proposition 4.4

In this appendix we provide a full proof of Proposition 4.4. We introduce a set up of the problem first and then restate the proposition for completeness.

In the following G is a group of fixed size n, and $\mathbb{K}$ is an infinite field with char $\mathbb{K} = 0$. We also denote by $S_d = \mathrm{Sym}_{\mathbb{K}}(n, d)$ the set of homogeneous polynomials of degree d in n variables, and fix $d = r^{L-1}$ where $r \geq 2$ and $L \geq 1$. For simplicity too, for filters $\theta = (\theta_1, \dots, \theta_L)$ we denote by $p_i(\theta) := \sigma_r(\Phi_\theta)(g_i) \in S_{rd}$ and the generated affine secant line by $W_\theta = \langle p_1(\theta), \dots, p_n(\theta) \rangle$. We also define $\theta^o := (e, \dots, e)$ to be the tuple of L identity filters, then $p_i(\theta^o) = x_i^{rd}$ and $W_{\theta^o} = \langle x_1^{rd}, \dots, x_n^{rd} \rangle$. Finally, all dimensions are Krull dimensions of $\mathbb{K}$-schemes and $\mathrm{length}(Y_\theta) = \dim_{\kappa(\theta)} \Gamma(Y_\theta, \mathcal{O}_{Y_\theta})$, where $\kappa(\theta)$ is the residue field at $\theta \in \Theta$.

Proposition G.1 (Proposition 4.4). *For general filters $\theta = (\theta_1, \dots, \theta_L)$, we have*

$$Y_\theta := \mathbb{P}(W_\theta) \cap VP_{n,r,d} = \{[p_1(\theta)], \dots, [p_n(\theta)]\}.$$

The approach we adopt for this proof consists of two main steps. The first step includes showing that the lemma holds for the special filters θ^o; more precisely, we show that the intersection scheme $Y_{\theta^o} = \mathbb{P}(W_{\theta^o}) \cap VP_{n,r,d}$ consists of n reduced points. In the second step, we propagate this argument to general filters using semi continuity arguments.

Lemma G.2. *The morphism $\pi_r : \mathbb{P}(S_d) \to \mathbb{P}(S_{rd}), [q] \mapsto [q^r]$ is a closed immersion. Hence $VP_{n,r,d} := \pi_r(\mathbb{P}(S_d)) \cong \mathbb{P}(S_d)$ and consequently, for $q \neq 0 \in S_d$ we have*

$$T_{q^r}\widehat{VP}_{n,r,d} = q^{r-1}S_d,$$

where $\widehat{VP}_{n,r,d}$ is the affine cone associated with the projective variety $VP_{n,r,d}$.

Proof. We only show that π_r is a closed immersion; the rest is a direct result. The property of being a closed immersion is fpqc local on the base (Stacks project authors, 2026, Lemma 02L6); which means given a faithfully flat quasi compact (fpqc) covering of the target, the morphism is a closed immersion if and only if the based-change morphism under each element in the cover is a closed immersion. The morphism $\mathbb{P}(S_{rd})_{\overline{\mathbb{K}}} \to \mathbb{P}(S_{rd})_{\mathbb{K}}$ obtained by base change from $\operatorname{Spec}\overline{\mathbb{K}} \to \operatorname{Spec}\mathbb{K}$ is faithfully flat and affine, hence a one-element fpqc covering of $\mathbb{P}(S_{rd})_{\mathbb{K}}$, and the base change of π_r along it is $\pi_{r,\overline{\mathbb{K}}}$. It therefore suffices to treat the case $\mathbb{K} = \overline{\mathbb{K}}$, which we assume for the rest of the proof.

To show π_r is a closed immersion, under the assumption $\mathbb{K} = \overline{\mathbb{K}}$, we rely on a theorem that says a scheme-morphism is a closed immersion if and only if it is proper, universally injective and unramified (Stacks project authors, 2026, Lemma 04XV). Next we prove each one of these properties in order.

π_r is proper. $\mathbb{P}(S_d)$ is proper over $\mathbb{K}$, $\mathbb{P}(S_{rd})$ is separated over $\mathbb{K}$, and π_r is a morphism over $\mathbb{K}$; hence π_r is proper by (Stacks project authors, 2026, Lemma 01W6).

π_r is universally injective. By (Stacks project authors, 2026, Lemma 01S4) the morphism π_r is universally injective if the extension $\pi_r : \mathbb{P}(S_d)(L) \to \mathbb{P}(S_{rd})(L)$ is injective where $L \supset \mathbb{K}$ is a field extension. Let $q_1, q_2 \in S_d(L) \subset L[x_g : g \in G]$ such that $q_1^r = \lambda q_2^r$ where $\lambda \in L^*$. Choose $\mu \in \overline{L}$ such that $\mu^r = \lambda$ then $q_1^r = (\mu q_2)^r$ in the unique factorization domain $\overline{L}[x_g : g \in G]$. Hence $q_1 = cq_2$ where $c \in \overline{L}^*$. However $q_1 \in L[x_g : g \in G]$ therefore by comparing nonzero coefficients of q_1 and cq_2 we conclude that $c \in L^*$. Hence, π_r is injective, and since L is an arbitrary field extension, it is universally injective.

π_r is unramified. At a closed point $[q] \in \mathbb{P}(S_d)$ we have the differential of π_r to be the map

$$S_d/\mathbb{K}q \to S_{rd}/\mathbb{K}q^r, \quad v + \mathbb{K}q \mapsto rq^{r-1}v + \mathbb{K}q^r.$$

The differential is injective because if $rq^{r-1}v = \mu q^r$, where $\mu \in \mathbb{K}$, then $q^{r-1}(rv - \mu q) = 0$, and since $S = \mathbb{K}[x_g : g \in G]$ is a domain then $rv = \mu q$. Hence $v \in \mathbb{K}q$, which makes the differential map injective. Since $\mathbb{K}$ is algebraically closed then π_r is unramified at the closed point q. Since the ramified locus is closed in $\mathbb{P}(S_d)$, which is $\mathbb{K}$-scheme of finite type. Therefore if the ramified locus is non-empty then it must include a closed point but since all closed points have been excluded then then the locus is empty and so π_r is unramified. □

Lemma G.3. *For the special filters θ^o we have $Y_{\theta^o} = \{[p_1(\theta^o)], \dots, [p_n(\theta^o)]\}$, as sets. Moreover, the tangent space of the intersection scheme Y_{θ^o} vanishes at all points, and so Y_{θ^o} consists of n reduced points and $\operatorname{length}(Y_{\theta^o}) = n$.*

Proof. Let $L \supset \mathbb{K}$ be a field extension, and let $[q] \in \mathbb{P}(W_{\theta^o})(L)$ then q is a sum of monomials where each one is is a power of a single variable (there are no monomials of mixed variables such as x_1x_2). Assume $[q] \in VP_{n,r,d}(L)$ then there is $t \in S_d \otimes_{\mathbb{K}} L$ such that $t^r = \lambda q$ for $\lambda \in L^*$. Note that we worked over the extension L just to show that field extension does not add any point to Y_{θ^o}.

If t is a not power of single variable then t^r contains monomial of mixed variables, due to char $\mathbb{K} = 0$, which contradicts our first observation. Hence t is a pure power and so is t^r and q. Hence $Y_{\theta^o} = \{[x_1^{rd}], \dots, [x_n^{rd}]\}$.

We compute tangent space at $p \in Y_{\theta^o}$ using the formula:

$$T_p Y_{\theta^o} = T_p \mathbb{P}(W_{\theta^o}) \cap T_p VP_{n,r,d}.$$

For $p = [x_i^{rd}]$, we have $T_{x_i^{rd}} W_{\theta^o} = W_{\theta^o}$, and by Lemma G.2 we have $T_{x_i^{rd}} \widehat{VP}_{n,r,d} = x_i^{d(r-1)} S_d$. At the intersection of these tangent spaces we get the polynomial equation

$$\sum_{j=1}^{n} c_j x_j^{rd} = x_i^{d(r-1)} q,$$

where $c_j \in \mathbb{K}$ and $q \in S_d$. Since $r > 1$, then setting $x_i = 0$ gives us

$$\sum_{j=1, j\neq i}^{n} c_j x_j^{rd} = 0$$

which implies $c_j = 0$ for all $j \neq i$. Therefore, $T_{x_i^{rd}} \widehat{VP}_{n,r,d} \cap T_{x_i^{rd}} W_{\theta^o} =< x_i^{rd} >$ which is killed by projectivization. Hence the intersection of the tangent spaces of the corresponding projective varieties is trivial.

Recall that by definition, the tangent space $T_{[x_i^{rd}]} Y_{\theta^o}$ is the dual to $\mathfrak{m}_{[x_i^{rd}]}/\mathfrak{m}^2_{[x_i^{rd}]}$ where $\mathfrak{m}_{[x_i^{rd}]}$ is the maximal ideal in the local ring $\mathcal{O}_{Y_{\theta^o}, [x_i^{rd}]}$. By the above $\mathfrak{m}_{[x_i^{rd}]}/\mathfrak{m}^2_{[x_i^{rd}]} = 0$, and by Nakayama Lemma, we get $\mathfrak{m}_{[x_i^{rd}]} = (0)$ is the zero ideal, hence the local ring $\mathcal{O}_{Y_{\theta^o}, [x_i^{rd}]}$ is a field and therefore $[x_i^{rd}]$ is a reduced point. Each $[x_i^{rd}]$ is a $\mathbb{K}$-rational point, so this field is $\kappa([x_i^{rd}]) = \mathbb{K}$. Hence

$$\Gamma(Y_{\theta^o}, \mathcal{O}_{Y_{\theta^o}}) \cong \mathbb{K}^n \qquad \text{and} \qquad \text{length}(Y_{\theta^o}) = n.$$

□

Lemma G.4. *Let $\Theta^o \subset \Theta$ be the locus where $p_1(\theta), \ldots, p_n(\theta)$ are linearly independent; note $\theta^o \in \Theta^o$. Then there is a dense open set $U \subset \Theta^o$ where for all $\theta \in U$ the intersection scheme Y_θ consists of n reduced points.*

Proof. In this proof, we define an incidence scheme Y and a projection morphism $\pi : Y \to \Theta^o$, such that the intersection scheme Y_θ coincide with the fiber scheme of π over $\theta \in \Theta^o$. In the first step, we show that π is proper. Then in two consecutive steps, relying on lemmata (Stacks project authors, 2026, Lemma 0D4I) and (Stacks project authors, 2026, Lemma 07ZC), we show that the subset in Θ^o where the intersection scheme Y_θ is of dimension 0 and has length n is open and dense in $\Theta = \mathbb{K}^{nL}$.

Consider the incidence scheme

$$Y := \{(\theta, [F]) \in \Theta^o \times VP_{n,r,d} \mid p_1(\theta) \wedge \ldots p_n(\theta) \wedge F = 0 \in \bigwedge^{n+1} S_{rd}\}.$$

The scheme Y is closed in $\Theta^o \times VP_{n,r,d}$ because it is defined by equations that are polynomials in θ and linear in F. This implies the inclusion $\iota : Y \hookrightarrow \Theta^o \times VP_{n,r,d}$ is a closed immersion and hence proper. Since $VP_{n,r,d}$ is projective then the projection pr_1 to Θ^o is proper. The composition of proper morphisms is proper, therefore the projection $\pi = pr_1 \circ \iota : Y \to \Theta^o$ is proper.

For $\theta \in \Theta^o$ the fiber scheme of π is the intersection scheme $Y_\theta := \mathbb{P}(W_\theta) \cap VP_{n,r,d}$. By Lemma G.3 we have Y_{θ^o} is of finite length equal to n. By (Stacks project authors, 2026, Lemma 0D4I) which states that for a proper morphism, the fiber dimension function over the base is upper semi-continuous. This implies the subset $B := \{\theta \in \Theta^o \mid \dim Y_\theta \geq 1\}$ is closed in Y. Then $U_1 := \Theta^o \backslash B$ open and non-empty because Y_{θ^o} is finite by Lemma G.3 and therefore $\theta^o \in U_1$. Our next goal is to shrink U_1 further so that it only contains parameters where their fibers are finite and of lenght n.

The morphism π is proper and quasi-finite over U_1 therefore it is finite by (Stacks project authors, 2026, Lemma 02LS). Finiteness of π implies the pushforward sheaf $\mathcal{F} = \pi_* \mathcal{O}_Y$ restricted over U_1 is coherent. Recall that the fiber $\mathcal{F} \otimes \kappa(\theta)$ of the sheaf $\mathcal{F}$ over the point θ is the $\kappa(\theta)$-vector space obtained by evaluating the germs in the stalk $\mathcal{F}_\theta$ over the point θ. The rank of the fiber is

the dimension of the vector space. Moreover finiteness of π over U_1 implies it is affine there and therefore the pushforward π_* commutes with an arbitrary base change (Stacks project authors, 2026, Lemma 02KG); in particular it commutes with the base change given by the morphism $\operatorname{Spec}\kappa(\theta) \to U_1$ that corresponds to the inclusion of a parameter θ in U_1. This commutativity implies $\mathcal{F} \otimes \kappa(\theta) \cong \dim_{\kappa(\theta)} \Gamma(Y_\theta, O_Y)$. Hence

$$\operatorname{rank}(\mathcal{F} \otimes \kappa(\theta)) = \dim_{\kappa(\theta)} \Gamma(\theta \times Y_\theta, O_Y) = \text{length } (Y_\theta).$$

The rank of a coherent sheaf is upper semi-continuous by (Stacks project authors, 2026, Lemma 07ZC), therefore there is a non-empty open set $\theta^o \in U_2 \subset U_1$ where $\text{length}(Y_\theta) \leq \text{length}(Y_{\theta^o}) = n$ for all $\theta \in U_2$. However, $\text{length}(Y_\theta) \geq n$ because it always includes $\{ p_1(\theta), \ldots, p_n(\theta)\}$ which are linearly independent by definition and therefore distinct. Hence, $\text{length}(Y_\theta) = n$ for all $\theta \in U_2$ which is open and hence dense in Θ^o, which itself is open and dense in $\Theta = \mathbb{K}^{nL}$, and therefore U_2 is open and dense in Θ as required. □

Proof of Proposition 4.8. Follows directly from Lemma G.4. □

H PROOFS FOR GENERAL FIBERS

In this appendix, we provide complete proofs for the theorems in Subsections §4.2 and §4.3.

H.1 GENERAL FIBER OF $\tilde{\varphi}$

Theorem (4.3)**.** *For activation degree $r \geq 2$ and general filters $\theta := (\theta_1, \ldots, \theta_L)$, we have $\tilde{\varphi}_\theta = \tilde{\varphi}_\psi$ if and only if*

$$\psi = (\theta_1 * g_1,\ g_1^{-1} * \theta_2 * g_2,\ \ldots,\ g_{L-2}^{-1} * \theta_{L-1} * g_{L-1},\ g_{L-1}^{-1} * \theta_L),$$

where $g_i \in G$ for $1 \leq i \leq L-1$, up to rescaling each filter.

Proof. We do a proof by induction over the number of layers L. The base case is trivial and the fiber is a , up to rescaling the only filter.

Assume the proposition holds for all numbers of layers less than or equal to $L-1$. Let θ' and ψ' denote the first $L-1$ filters in θ and ψ, and consider the case for L to get

$$\varphi_\theta = \varphi_{\theta'}^{\otimes r} * \theta_L = \lambda \varphi_{\psi'}^{\otimes r} * \psi_L = \lambda \varphi_\psi,$$

where $\lambda \in \mathbb{K}\backslash\{0\}$. Since θ is general, by Proposition D.5 we get that $\varphi_\theta \in \mathbb{K}[G^{r^{L-1}}]$ has an inverse filter φ_θ^{-1} and, moreover, $\psi_L \in \mathbb{K}[G^{r^{L-1}}]$ has an inverse given by $\psi_L^{-1} = \varphi_\theta^{-1} * \lambda \varphi_{\psi'}^{\otimes r}$. Convolute both sides above by ψ_L^{-1} to get

$$\varphi_{\theta'}^{\otimes r} * \theta_L * \psi_L^{-1} = \lambda \varphi_{\psi'}^{\otimes r}.$$

Define $\mu := \theta_L * \psi_L^{-1}$ and substitute in the above equation to get

$$\varphi_{\theta'}^{\otimes r} * \mu = \lambda \varphi_{\psi'}^{\otimes r}.$$

Write $\mu = \sum_{g \in G} \mu(g) g$ and use Proposition 2.6 to distribute convolution over the Kronecker product to get

$$\sum_{g \in G} \mu(g) \varphi_{\theta'} * g \otimes \ldots \otimes \varphi_{\theta'} * g = \lambda \varphi_{\psi'} \otimes \ldots \otimes \varphi_{\psi'}.$$

Notice that for any $g \in G$, $\varphi_{\theta'} * g$ is a vector in $\mathbb{K}^{n^{r^{L-2}}}$, and more importantly, it is a column in the matrix $\text{Mat}_{\varphi_{\theta'}}$. For general filters $\theta_1, \ldots, \theta_{L-1}$, Proposition D.5 shows the determinant $\det(\text{Mat}_{\varphi_{\theta'}}) \neq 0$ and so the vectors $\{\varphi_{\theta'} * g \mid g \in G\}$ are linearly independent. Therefore the vectors $\{\mu(g)^{1/r} \varphi_{\theta'} * g \mid g \in G, \mu(g)^{1/r} \neq 0\}$ are linearly independent in $\overline{\mathbb{K}}^{n^{r^{L-2}}}$. By Lemma 4.2, the above equation holds if and only if $\mu \in G$ up to a scaling, which implies $\theta_L = g * \psi_L$ up to a rescaling. The rest follows by induction. □

H.2 General fiber of Φ

Theorem (4.5). *For activation degree $r \geq 2$ and general filters $\theta := (\theta_1, \ldots, \theta_L)$, we have $\tilde{\Phi}_\theta = \tilde{\Phi}_\psi$ if and only if $\psi := (\theta_1 * g_1,\ g_1^{-1} * \theta_2 * g_2,\ \ldots,\ g_{L-2}^{-1} * \theta_{L-1} * g_{L-1},\ g_{L-1}^{-1} * \theta_L)$, where $g_i \in G$ for $1 \leq i \leq L-1$, up to rescaling each filter.*

Proof. The proof is similar to Theorem 4.3 and proceeds by induction over the number of layers L. The base case is trivial and the fiber is a singleton, up to rescaling.

Assume the theorem holds for all numbers of layers less than or equal to $L-1$. Let θ' and ψ' denote the first $L-1$ filters in θ and ψ, and consider the case for L to get

$$\Phi_\theta = \sigma_r(\Phi_{\theta'}) * \theta_L = \lambda \sigma_r(\Phi_{\psi'}) * \psi_L = \lambda \Phi_\psi, \tag{13}$$

where $\lambda \in \mathbb{K} - \{0\}$. We show ψ_L is invertible in $\mathbb{K}[G]$. Let $S = K[x_g \mid g \in G]$ be the polynomial ring over $\mathbb{K}$ and $\mathrm{Frac}(S)$ be its field of fraction. Since θ is general by Proposition D.4 $\det \Phi_\theta \neq 0$ in $S[G]$ and therefore Φ_θ is invertible in $\mathrm{Frac}(S)[G]$. Then Eq (13) implies ψ_L has a left inverse in $\mathrm{Frac}(S)[G]$ which is a finite dimensional algebra and therefore ψ_L has a right inverse, too, and it is invertible in $\mathrm{Frac}(S)[G]$. However, $\psi_L \in \mathbb{K}[G] \subset \mathrm{Frac}(S)[G]$. Therefore ψ_L inverse in $\mathbb{K}[G]$ exists.

Define $\mu := \theta_L * \psi_L^{-1}$ and substitute in the last equation to get

$$\sigma_r(\Phi_{\theta'}) * \mu = \lambda \sigma_r(\Phi_{\psi'}).$$

Then evaluate at $e \in G$ to get

$$\sum_{g \in G} \sigma_r(\Phi_{\theta'})(g) \mu(g^{-1}) = \lambda \sigma_r(\Phi_{\psi'})(e).$$

The last equation says that $\lambda \sigma_r(\Phi_{\psi'})(e)$ lies in the n-secant line generated by $\{\sigma_r(\Phi_{\theta'})(g)\}_{g \in G}$. By Proposition 4.4, we get $\lambda \sigma_r(\Phi_{\psi'})(e) = \sigma_r(\Phi_{\theta'})(g)$ for some $g \in G$. As a result, we have $\mu = \lambda g$ for some $g \in G$, or equivalently $\theta_L = \lambda g * \psi_L$. The rest follows by induction.

□

I NUMERICAL VERIFICATION

In this appendix, we iterate the cases we are able to numerically check our dimension formula (Theorem 4.1) and the size of general fibers (Theorems 4.3, 4.5) by developing a package based on `Sage` and `Macaulay2`.

For the dimension, we perform sanity checks for small networks (groups with up to 13 elements and network depth up to 3). However, the numerical computation becomes intractable for bigger groups and deeper network because the size of the Jacobian matrix, for both maps φ and Φ, is at least quadratic in terms of the size of the group and exponential in terms of the number of layers.

For fiber size computation, we perform sanity checks for groups and networks as described in Table 1. The verification for bigger examples is intractable since the algorithm to compute the fiber's size requires to saturate with respect to the irrelevant ideal which is very expensive.

Group	$\mathbb{Z}_2$	$\mathbb{Z}_3$	$\mathbb{Z}_4$	$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\mathbb{Z}_5$	$\mathbb{Z}_6$	$\mathbb{Z}_2 \times \mathbb{Z}_3$	S_3
Max no. of layers	4	3	2	2	2	2	2	2

Table 1: Numerical verification of Theorem 4.5 for PGCNN with the listed groups and maximum depth as shown in the second row; the activation degree is fixed to 2.