The Geometry of Polynomial Group Convolutional Neural Networks

Yacoub Hendi
Uppsala University

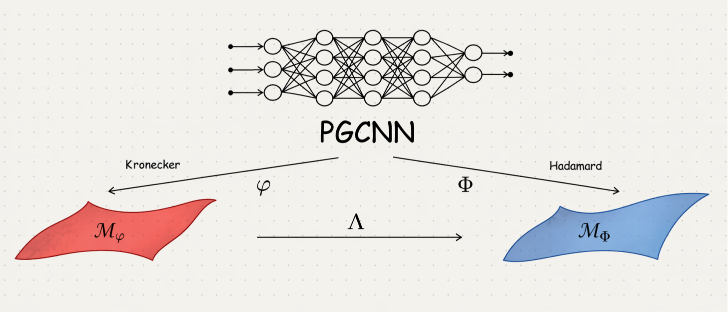
Based on a joint work with Daniel Persson (Chalmers) and Magdalena Larfors (UU)
(arXiv: 2603.29566)

May 2026

UPPSALA UNIVERSITET

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

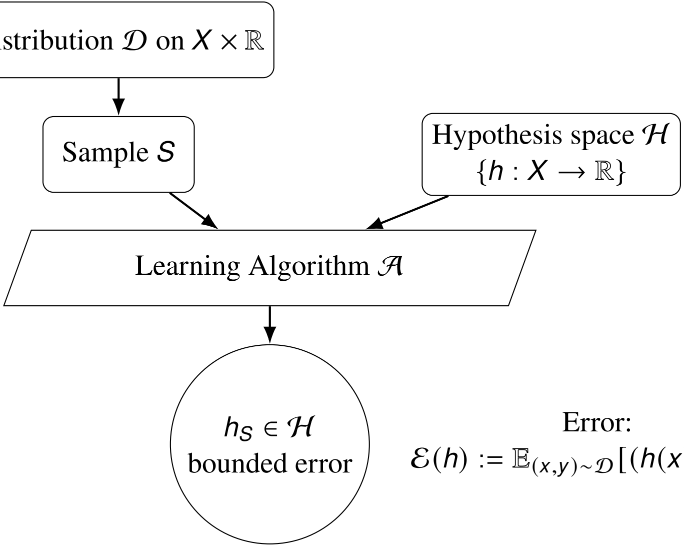
Outline



- Introduction to Learning Theory
- Neuroalgebraic Geometry:
  from Learning Theory to Algebraic Geometry
- Polynomial Neural Networks
- Polynomial Group Convolutional Neural Networks

Learning Theory & Neuroalgebraic Geometry

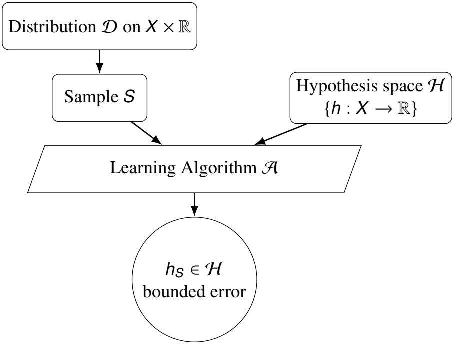
Learning Theory



Error:

$$\mathcal{E}(h) := \mathbb{E}_{(x,y)\sim\mathcal{D}}[(h(x) - y)^2].$$

Learning Theory



- **Goal**: minimizes

$$\mathcal{E}(h) := \mathbb{E}_{(x,y)\sim\mathcal{D}}[(h(x) - y)^2].$$

- **Method** (Empirical Risk Minimization $\mathcal{A}$): minimize

$$\mathcal{E}_S(h) := \frac{1}{m} \sum_{(x,y)\in S} (h(x) - y)^2.$$

Learning Theory

- **Question**: Is it possible to *probabilistically* bound *the generalization error* for all $h \in \mathcal{H}$:
$$|\mathcal{E}(h) - \mathcal{E}_S(h)|.$$

Probable Approximate Correct Learnability

Definition (PAC Learnability)

A hypothesis class $\mathcal{H}$ is **PAC learnable** if there exist a learning rule $\mathcal{A}$ and a sample complexity function:

$$m_{\mathcal{H}} : (0,1)^2 \longrightarrow \mathbb{N},$$

such that for every $\varepsilon, \delta \in (0,1)$ and every distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$:

If

$$m \geq m_{\mathcal{H}}(\varepsilon, \delta)$$

i.i.d. samples from $\mathcal{D}$, then with probability at least $1 - \delta$, $\mathcal{A}$ returns a hypothesis $h_{\mathcal{A}}$ such that

$$\mathcal{E}(h_{\mathcal{A}}) - \min_{h \in \mathcal{H}} \mathcal{E}(h) \leq \varepsilon.$$

Sample Complexity & Covering numbers

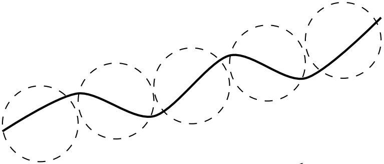
- $\mathcal{H} \subset C(X)$ – compact set, d – metric on $C(X)$.

$$m_{\mathcal{H}}(\varepsilon, \delta) = \Omega\left(\frac{A}{\varepsilon^2} \log \frac{\mathcal{N}(\mathcal{H}, \frac{\varepsilon}{B})}{\delta}\right).$$

$\mathcal{N}(\mathcal{H}, \varepsilon)$ = minimum number of ε-balls needed to cover $\mathcal{H}$.

[Marchetti et al., 2025, Cucker and Smale, 2001]

Sample Complexity & Covering numbers



ε-balls covering $\mathcal{H}$

$\mathcal{H}$–algebraic variety

$$\log \mathcal{N}(\mathcal{H}, \varepsilon) = \mathcal{O}\left(\dim(\mathcal{H}) \log \frac{\deg(\mathcal{H})}{\varepsilon} + A\right).$$

[Marchetti et al., 2025, Yomdin and Comte, 2004]

Take away

Dimension and degree $\implies$ covering numbers $\implies$ expressivity and sample complexity of the model

Neuroalgebraic Geometry

Machine Learning	Algebraic Geometry
sample complexity and expressivity	dimension, degree, and covering number
subnetworks and implicit bias	singularities
identifiability and invariance	fibers of the parameterization
optimization and gradient descent	critical point theory, discriminants, and dynamical invariants

[Marchetti et al., 2025]

Polynomial Neural Networks

Definition of Polynomial Neural Networks

- A L-layer polynomial neural networks with
  - input layer x
  - weight matrices $\theta = (W_1, \dots, W_L) \in \mathbb{R}^N$
  - monomial activation function σ_r

$$\Phi_\theta(x) := W_L \sigma_r(W_{L-1}\sigma_r(\dots \sigma_r(W_2 \sigma_r(W_1 x)))).$$

- $\Phi(\theta, x) \in \mathrm{Sym}(x, r^{L-1})^{N_L}$, coefficients $\in \mathrm{mSym}(\theta, (r^{L-1}, r^{L-2}, \dots, 1))$.

- **Example:** $L = 2, r = 2$. Let $x = (x_1, x_2)$,

$$W_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad W_2 = \begin{pmatrix} e & f \end{pmatrix}.$$

Then

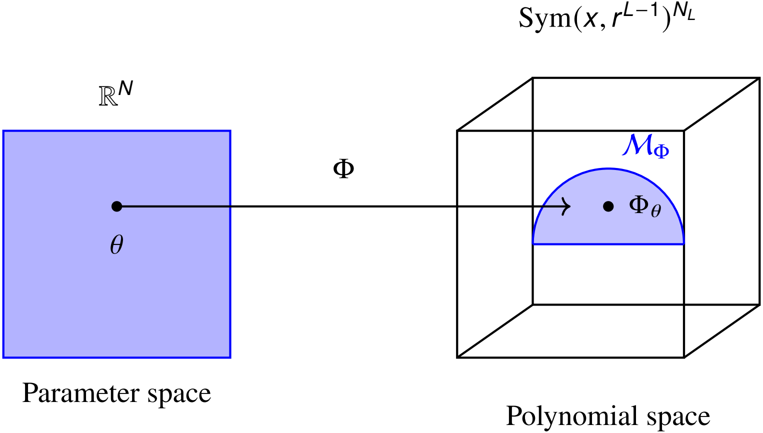
$$\Phi_\theta(x) = W_2\,\sigma_2(W_1 x) = e(ax_1 + bx_2)^2 + f(cx_1 + dx_2)^2.$$

Definition of Neuromanifold

- The neuromanifold $\mathcal{M}_\Phi$ is the image of Φ

$$\mathcal{M}_\Phi := \{\Phi_\theta \| \theta \in \mathbb{R}^N\} \subset \mathrm{Sym}(x, r^{L-1})^{N_L}.$$

By Tarski-Seidenberg, $\mathcal{M}_\Phi$ is a semi-algebraic set.



Related works

- Fully connected networks; no constraints of W_i.
- For high enough activation degree, the neuromanifold has the expected dimension. [Craciun, 2025, Finkel et al., 2025, Massarenti and Mella, 2025, Usevich et al., 2025, Kileel et al., 2019].
- Convolutional networks (Shift Equivariant); W is cyclic.

$$W = \begin{pmatrix} w_1 & w_2 & w_3 & 0 & 0 \\ 0 & w_1 & w_2 & w_3 & 0 \\ 0 & 0 & w_1 & w_2 & w_3 \end{pmatrix}.$$

The neuromanifold is a variety that is birational a Segre-Veronese Variety. [Shahverdi et al., 2024].

Polynomial Group Convolutional Neural Networks

Group algebras

Let G-finite group, $\mathbb{K}$-field, x-tuple of variables, $S := \mathbb{K}[x]$.

- **Group algebra:**

$$S[G] = \{\theta : G \to S\} \cong \left\{ \sum_{g \in G} \theta(g) g \mid \theta(g) \in S \right\} \cong S^{|G|}.$$

- **Convolution**: $(\theta * \psi)(g) = \sum_{h \in G} \theta(gh^{-1})\psi(h)$.
- **Hadamard (entrywise):** $(\theta \circ \psi)(g) = \theta(g)\psi(g)$, $\sigma_r(\theta) = \underbrace{\theta \circ \cdots \circ \theta}_{r}$.
- **Kronecker:** $\theta \otimes \psi \in S[G \times H]$, $(\theta \otimes \psi)(g, h) = \theta(g)\psi(h)$,

Definition of PGCNN

Let $x = x_1 g_1 + \cdots + x_n g_n \in S[G]$ and $\theta = (\theta_1, \ldots, \theta_L) \in (\mathbb{K}[G])^L$.

- **PGCNN (Hadamard view):**

$$\Phi_\theta(x) = \sigma_r(\ldots(\sigma_r(x * \theta_1) * \theta_2) \cdots * \theta_{L-1}) * \theta_L,$$

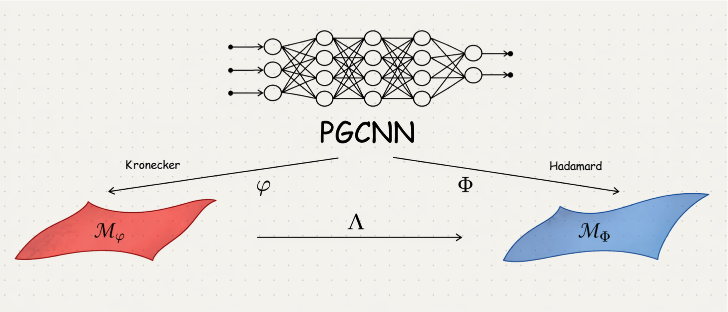
Use $\sigma_r(x * \theta_1) = (x^{\otimes r} * \theta_1^{\otimes r})|_G$ to get

- **Kronecker view:**

$$\Phi_\theta = (x^{\otimes r^{L-1}} * \underbrace{\theta_1^{\otimes r^{L-1}} * \theta_2^{\otimes r^{L-2}} * \cdots * \theta_L}_{\varphi_\theta})\Big|_G$$

- **Neuromanifold (for $r \geq 2$):**
  $\mathcal{M}_\Phi = \{\Phi_\theta : \theta \in \mathbb{K}[G]^L\} \subset \mathrm{Sym}(x, r^{L-1})[\mathrm{G}]$,
  $\mathcal{M}_\varphi = \{\varphi_\theta : \theta \in \mathbb{K}[G]^L\} \subset \mathbb{K}[G^{r^{L-1}}]$.

Definition of PGCNN



$$\mathbb{K}[G]^L$$

$$\varphi \swarrow \qquad \searrow \Phi$$

$$\mathcal{M}_\varphi \subset \mathbb{K}[G^{r^{L-1}}] \xrightarrow{\Lambda} \mathcal{M}_\Phi \subset \mathrm{Sym}(x, r^{L-1})[G]$$

Main results: Dimension of neuromanifold

Theorem

Let $L = \#$layers, and $r \geq 2$ then

$$\dim(\mathcal{M}_\varphi) = \dim(\mathcal{M}_\Phi) = L(|G| - 1) + 1.$$

Proof techniques:

- At a general point $\theta \in \mathbb{K}[G]^L$, $\operatorname{rank}(J_\theta \Phi) = \dim(\mathcal{M}_\Phi)$.
- Consider $J_\theta \Phi$ as a matrix over the ring of polynomials $\mathbb{K}[\theta]$. Compute the symbolic rank.
- **Example:** 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

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

Main results: shape of general fiber of φ

Theorem

For general filters $\theta := (\theta_1, \ldots, \theta_L)$, we have $\varphi_\theta = \varphi_\psi$, up to a rescaling, 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 $1 \le i \le L-1$, up to rescaling each filter.

Lemma (Lemma 5.1 [Comon et al., 2008])

Let $y_1, \ldots, y_n \in \mathbb{K}^d$ be linearly independent. Then the symmetric tensor

$$A := \sum_{i=1}^{n} y_i^{\otimes r}$$

has a symmetric rank $\mathrm{rank}_S(A) = n$.

General Fiber of φ: Key Proof Steps

Step 1: Separate the last layer. Write $\theta' = (\theta_1, \ldots, \theta_{L-1})$. From

$$\varphi(\theta) = \lambda\varphi(\psi)$$

we obtain

$$\varphi(\theta')^{\otimes r} * \theta_L = \lambda\, \varphi(\psi')^{\otimes r} * \psi_L.$$

Step 2: Invert the last layer (generality). For general filters θ, ψ_L is invertible. Set

$$\mu := \theta_L * \psi_L^{-1}, \quad \text{so} \quad \varphi(\theta')^{\otimes r} * \mu = \lambda\, \varphi(\psi')^{\otimes r}.$$

Step 3: Use symmetric tensor rank. This identity expands as

$$\sum_{g \in G} \mu(g) \left(\varphi(\theta') * g\right)^{\otimes r}.$$

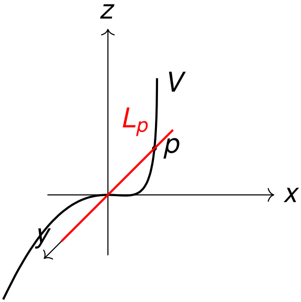
- The vectors $\{\varphi(\theta') * g\}_{g \in G}$ are linearly independent.
- By a classical rank result [Comon et al., 2008], this forces μ to be a (scaled) group element.

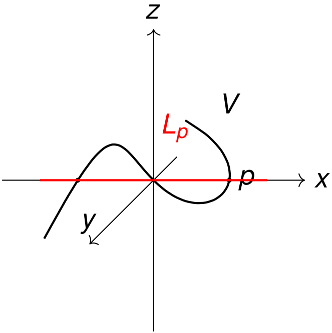
Conclusion. θ_L and ψ_L differ by group action and scale. Apply the inductive hypothesis to θ'.

Main results: shape of general fiber of Φ

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

Choose m polynomials in $p_1, \ldots, p_m \in VP_{n,k,d}$, we call their linear span $< p_1, \ldots, p_m >$ in $\mathrm{Sym}(x, kd)$ by an m-*secant* line of $VP_{n,k,d}$.





Main results: general fiber of Φ

Conjecture

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

$$< \sigma_r(\Phi_\theta)(g_1), \ldots, \sigma_r(\Phi_\theta)(g_n) > \cap VP_{n,r,r^{L-1}} = \{\sigma_r(\Phi_\theta)(g_1), \ldots, \sigma_r(\Phi_\theta)(g_n)$$

Theorem (Conditional on Conjecture)

For general filters $\theta := (\theta_1, \ldots, \theta_L)$, *we have* $\Phi_\theta = \Phi_\psi$, *up to a rescaling, 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.*

Outlook

- Prove the conjecture about general fibers of Φ.
- Study the degree of $\mathcal{M}_\Phi$.
- Study the singular locus of the neuromanifold.
- Extend to arbitrary polynomial and rational activation functions.
- Extend to compact groups.

Thanks for listening!

References I

Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank, 2008. URL https://arxiv.org/abs/0802.1681.

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

Felipe Cucker and Steve Smale. On the mathematical foundations of learning. 2001. URL http://www.mit.edu/~9.520/Papers/cuckersmale.pdf.

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.

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.

References II

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.

Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. On the geometry and optimization of polynomial convolutional networks, 2024. URL https://arxiv.org/abs/2410.00722.

Konstantin Usevich, Ricardo Borsoi, Clara Dérand, and Marianne Clausel. Identifiability of deep polynomial neural networks, 2025. URL https://arxiv.org/abs/2506.17093.

Yosef Yomdin and Georges Comte. *Tame geometry with application in smooth analysis*, volume 1834 of *Lecture Notes in Mathematics*. Springer Science & Business Media, Berlin, Heidelberg, 2004. ISBN 978-3-540-20612-5. doi: 10.1007/978-3-540-40960-1.