Minimal Filling Architectures of Polynomial Neural Networks: Counterexamples, Frontier Search, and Defects

Kevin Dao [*1] **Jose Israel Rodriguez** [*1]

Abstract

We provide counterexamples to the unimodal minimal filling architecture conjecture for polynomial neural networks (PNNs) with power activation functions. Fixing the input and output widths, the conjecture states that any minimal filling architecture has unimodal widths for the hidden layers. We found counterexamples via a frontier search, recursive dimension bounds on neurovarieties, and symbolic computation. Notably, several subarchitectures of our main example exhibit large defect, in contrast with the predominantly small-defect behavior observed in prior literature.

1. The Unimodal Minimal Filling Architecture Conjecture

Neuroalgebraic geometry studies the geometry of function spaces parameterized by machine-learning models when they are algebraic or semi-algebraic in nature. It relates algebro-geometric invariants (e.g., dimension and singularities) to notions such as expressivity and training behavior (Marchetti et al., 2025). This viewpoint involves other geometry-driven approaches in learning theory, including tropical-geometry descriptions of ReLU networks (Zhang et al., 2018; Brandenburg et al., 2024). In this article, we focus on polynomial neural networks; related neuroalgebraic analyses of model dimension and identifiability have also been carried out for other polynomial architectures, for example lightning self-attention (Henry et al., 2025).

We state the unimodal minimal filling architecture conjecture in its original form, Conjecture 12 of (Kileel et al., 2019), and recall the necessary definitions.

Definition 1.1. A (feedforward) *polynomial neural network* (PNN) with activation function σ is a function of the form

$$p_\theta : \mathbb{R}^{d_0} \to \mathbb{R}^{d_L}, \quad p_\theta(x) = W_L \sigma W_{L-1} \sigma \dots \sigma W_1 x$$

where σ acts entrywise by $(\sigma(z))_j = z_j^r$ for some $r \in \mathbb{N}$. The parameters are $\theta = (W_1, \dots, W_L)$, with each *weight matrix* $W_i \in \mathbb{R}^{d_i \times d_{i-1}}$. The network's *architecture* is the sequence $\mathbf{d} = (d_0, \dots, d_L)$, and the network's *depth* is L.

It is immediate that p_θ is a homogeneous polynomial map $\mathbb{R}^{d_0} \to \mathbb{R}^{d_L}$ and that $p_\theta \in \mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0})^{\oplus d_L}$. Since homogeneous polynomials of degree $D = r^{L-1}$ correspond to symmetric order-D tensors, a depth-L PNN induces a multi-stage factorization of such tensors.

Remark 1.2. In this article we use the terminology "feedforward polynomial neural network" to be consistent with (Kubjas et al., 2024; Kileel et al., 2019). In the literature, these are also referred to as "monomial multilayer perceptrons (MLPs)". The MLP terminology is useful when distinguishing fully connected feed forward neural networks from convolutional neural networks. Moreover, prefacing MLP with monomial can clarify when the activation function is monomial, as opposed to a general polynomial in a single variable.

Definition 1.3. For fixed r and architecture $\mathbf{d}$, the map

$$\Psi_{\mathbf{d},r} : \mathbb{R}^{d_1 \times d_0} \times \dots \times \mathbb{R}^{d_L \times d_{L-1}} \to \mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0})^{\oplus d_L}$$

given by $\theta \mapsto \begin{bmatrix} p_{\theta 1} & \dots & p_{\theta d_L} \end{bmatrix}^T$ where each $p_{\theta i}$ is a polynomial in the entries of the weight matrices is called the *parameter map*. The codomain is referred to as the *ambient space*, and its dimension is well known:

$$\dim\left(\mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0})^{\oplus d_L}\right) = d_L \binom{d_0 + r^{L-1} - 1}{r^{L-1}}.$$

The image of $\Psi_{\mathbf{d},r}$ is a set of vectors of polynomials denoted $\mathcal{F}_{\mathbf{d},r}$ which we call the *functional space*. Its Zariski closure $\mathcal{V}_{\mathbf{d},r} = \overline{\mathcal{F}_{\mathbf{d},r}}$ is called the *neurovariety* as it is an irreducible algebraic variety. To simplify notation, we remove r from the subscript when the context is clear.

One goal of neuroalgebraic geometry is to characterize when the neurovariety and ambient space dimension agree.

[1]Department of Mathematics, University of Wisconsin-Madison, Wisconsin, USA. Correspondence to: Kevin Dao <ktdao@wisc.edu>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

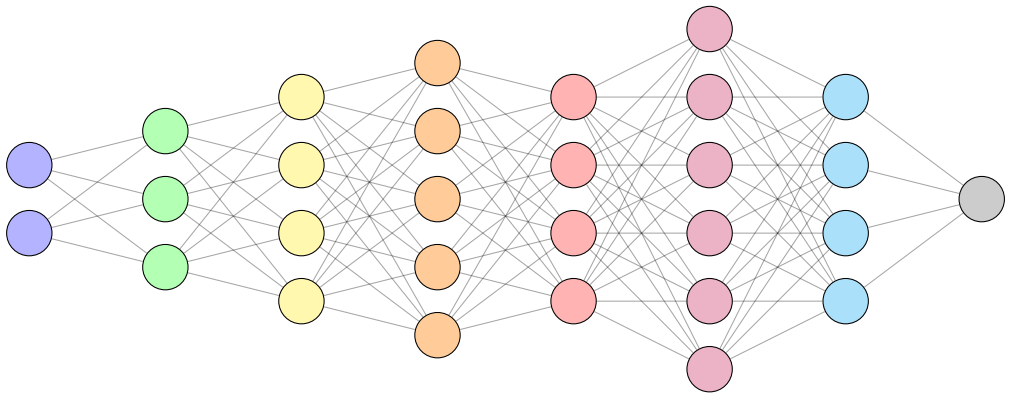

Figure 1. Our first counterexample to the unimodal minimal filling architecture conjecture ($r = 2$)

Example 1.4. Let $\mathbf{d} = (2, d_1, 1)$, and denote the input variable into the network by $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. If the weights are

$$W_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & d_1 \end{bmatrix} \quad \text{and} \quad W_2 = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix},$$

then $\Psi_{(2,d_1,1),r}(W_1, W_2)$ is the polynomial function

$$p_{W_1,W_2}(x) = (x_1+x_2)^r+(x_1+2x_2)^r+\cdots+(x_1+d_1x_2)^r$$

in $\mathrm{Sym}_r(\mathbb{R}^2)$, i.e., an order-$r$ symmetric tensor in $(\mathbb{R}^2)^{\otimes r}$.

One reason to focus on the neurovariety is due to Proposition 5 of (Kileel et al., 2019) which relates the dimension of a neurovariety $\mathcal{V}_{(d_0,\mathbf{a},d_L),r}$ with the expressivity of the functional space $\mathcal{F}_{(d_0,2\mathbf{a},d_L),r}$. By studying the neurovariety, one opens up the toolbox provided by algebraic geometry. A natural question is to study dimension (and other geometric properties) of the neurovariety. Dimension relates to the **expressivity** of the PNN.

The study of the dimension of neurovarieties has been done in various works such as (Kileel et al., 2019; Kubjas et al., 2024; Finkel et al., 2025; Usevich et al., 2025; Massarenti & Mella, 2025).

Definition 1.5. The neurovariety $\mathcal{V}_{\mathbf{d},r}$ is said to be *filling* if $\mathcal{V}_{\mathbf{d},r} = \mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0})^{\oplus d_L}$. If r is fixed, we will say that the architecture $\mathbf{d}$ is filling if its neurovariety is filling.

We define a partial ordering on architectures of the same depth L by declaring that $\mathbf{d}' \leq \mathbf{d}$ if $d'_i \leq d_i$ for $0 \leq i \leq L$. If $\mathbf{d}' < \mathbf{d}$, we refer to $\mathbf{d}'$ as a *subarchitecture* of $\mathbf{d}$.

Definition 1.6. For fixed r, an architecture $\mathbf{d} = (d_0, \dots, d_L)$ is *minimal filling* if $\mathbf{d}$ is filling and every subarchitecture $\mathbf{d}' = (d_0, d'_1, \dots, d'_{L-1}, d_L)$ with $\mathbf{d}' < \mathbf{d}$ is not filling. We say it is a *minimal filling architecture (MFA)*.

We are ready to state the unimodal minimal filling architecture conjecture of (Kileel et al., 2019).

Conjecture 1.7. Fix r, L, d_0, d_L. If $\mathbf{d} = (d_0, d_1, \dots, d_L)$ is a MFA, then there is an i with $0 \leq i \leq L$ such that

$$d_0 \leq d_1 \leq \cdots \leq d_i \qquad \& \qquad d_i \geq d_{i+1} \geq \cdots \geq d_L.$$

That is, the sequence $\mathbf{d}$ is **unimodal**.

The motivation for the conjecture is that it conforms with "conventional wisdom". In machine learning, the commonly held belief is that the network expands its capacity for feature representation to capture intricate patterns, and then narrows to refine these features for predictions. One reason to have believed this conjecture is that it holds in the shallow setting, where it follows from the Alexander–Hirschowitz theorem via the classical secant-Veronese dimension count.

2. The Counterexample: $(2, 3, 4, 5, 4, 6, 4, 1)$

Fix $r = 2$. Our main contribution addresses Conjecture 12 in (Kileel et al., 2019).

Theorem 2.1. *The architecture* $\mathbf{d} = (2, 3, 4, 5, 4, 6, 4, 1)$ *is a counterexample to the unimodal minimal filling architecture conjecture. That is, it is* ***minimal filling****, and* ***not unimodal****.*

The architecture $\mathbf{d} = (2, 3, 4, 5, 4, 6, 4, 1)$ is clearly not unimodal. We are tasked with showing it is minimal filling. Our approach is to prove (i) the neurovariety $\mathcal{V}_{\mathbf{d},2}$ is filling in Lemma 2.2 and (ii) no subarchitecture is filling in Lemma 2.6. Our proofs involve computing and upper-bounding the dimension of neurovarieties.

Lemma 2.2. *For* $\mathbf{d}$ *as in Theorem 2.1,* $\mathcal{V}_{\mathbf{d},2} = \mathrm{Sym}_{2^6}(\mathbb{R}^2) \cong \mathbb{R}^{65}$.

Proof. It suffices to show $\dim \mathcal{V}_{\mathbf{d},2} = 65$, the dimension of the ambient space. For more background, the reader can refer to Appendix A. We consider the Jacobian of the parameter map $\mathrm{Jac}\, \Psi_{\mathbf{d},r}$. By Theorem A.2,

$$\mathrm{rank}(\mathrm{Jac}\, \Psi_{\mathbf{d},r}(\theta)) \leq \dim \mathcal{V}_{\mathbf{d},r}$$

for every $\theta \in \mathbb{R}^{d_1 \times d_0} \times \cdots \times \mathbb{R}^{d_L \times d_{L-1}}$. By Corollary A.6 and the fact that $\Psi_{\mathbf{d},r}$ is given by polynomials with integer coefficients,

$$\begin{aligned}\operatorname{rank}(\operatorname{Jac} \Psi_{\mathbf{d},r}(\theta) \bmod p) &\leq \dim \mathcal{V}_{\mathbf{d},r} \\ &\leq \dim \operatorname{Sym}_{r^{L-1}}(\mathbb{R}^{d_0})^{\oplus d_L}\end{aligned}$$

where p is a prime and $\theta \in \mathbb{Z}^{d_1 \times d_0} \times \cdots \times \mathbb{Z}^{d_L \times d_{L-1}}$. Using the backpropagation algorithm as implemented in (Kubjas et al., 2024) we found[1] a prime p and θ for which

$$\operatorname{rank}(\operatorname{Jac} \Psi_{\mathbf{d},r}(\theta) \bmod p) = 65.$$

Since the ambient space also has dimension 65, we conclude $\dim \mathcal{V}_{\mathbf{d},r} = 65$. □

To show minimality, we utilize the following two lemmas.

Lemma 2.3. *(Kubjas et al., 2024) Fix $r \geq 1$. Let $\mathbf{d} = (d_0, \ldots, d_L)$ and $\mathbf{d}' = (d_0, \ldots, d_i', \ldots, d_L)$ be two architectures except they differ in exactly the ith entry for $0 < i < L$. Assume that $d_i' \leq d_i$. Then,*

$$\mathcal{F}_{\mathbf{d}'} \subseteq \mathcal{F}_{\mathbf{d}} \qquad \& \qquad \mathcal{V}_{\mathbf{d}'} \subseteq \mathcal{V}_{\mathbf{d}}.$$

In particular, $\dim(\mathcal{V}_{\mathbf{d}'}) \leq \dim(\mathcal{V}_{\mathbf{d}})$.

By induction, one can generalize Lemma 2.3 in the obvious manner: if $\mathbf{d}' \leq \mathbf{d}$ with $d_0' = d_0$ and $d_L' = d_L$, then $\dim(\mathcal{V}_{\mathbf{d}'}) \leq \dim(\mathcal{V}_{\mathbf{d}})$.

Lemma 2.4. *(Kileel et al., 2019) Fix $r \geq 1$ and let $k \in \{1, \ldots, L-1\}$. Then,*

$$\dim \mathcal{V}_{(d_0,\ldots,d_L)} \leq \dim \mathcal{V}_{(d_0,\ldots,d_k)} + \dim \mathcal{V}_{(d_k,\ldots,d_L)} - d_k.$$

By Lemma 2.3, $\mathbf{d}$ is minimal filling if the maximal subarchitectures obtained by decreasing a single hidden-layer width by one are not filling. These maximal subarchitectures are listed in Table 1. The suggested dimensions of these neurovarieties are obtained via finite field computations.

We will utilize the fillingness of certain architectures which are collected in the following proposition. Its proof follows exactly the same computational methods as Lemma 2.2.

Proposition 2.5. *For $r = 2$, the following architectures are filling:* $(2,4,5,4)$, $(2,3,3)$, $(2,3,4,4)$, $(2,3,4,5,3)$, $(3,6,4,1)$, $(2,3,4,5,4)$, *and* $(4,6,3)$.

The recursive bounds in Lemma 2.4 involve intermediate architectures whose filling status must be determined. Proposition 2.5 addresses the cases needed for the proof of the next lemma.

Lemma 2.6. *The ambient space is 65 dimensional. The subarchitectures $\mathbf{s}_i$ for $1 \leq i \leq 6$ as described in Table 1 have $\dim(\mathcal{V}_{\mathbf{s}_i,2})$ satisfying the bounds in the third column of the table.*

[1] Our choice of parameters and prime are in Appendix C.

Table 1. Bounds proved in Lemma 2.6 and reported $\dim \mathcal{V}_{\mathbf{s}_i,r}$.

ARCHITECTURE	REPORTED DIMENSION	PROVEN BOUND
$\mathbf{s}_1 = (2,2,4,5,4,6,4,1)$	35	≤ 53
$\mathbf{s}_2 = (2,3,3,5,4,6,4,1)$	60	≤ 61
$\mathbf{s}_3 = (2,3,4,4,4,6,4,1)$	62	≤ 63
$\mathbf{s}_4 = (2,3,4,5,3,6,4,1)$	39	$= 39$
$\mathbf{s}_5 = (2,3,4,5,4,5,4,1)$	61	≤ 62
$\mathbf{s}_6 = (2,3,4,5,4,6,3,1)$	59	≤ 62

Proof. Recall $r = 2$. For an architecture $\mathbf{d}$, let $D_{\mathbf{d}} = \dim(\mathcal{V}_{\mathbf{d}})$. Note that $D_{\mathbf{d}}$ is always bounded above by the dimension of the ambient space $D_{\mathbf{d}} \leq d_L \binom{d_0 + r^{L-1} - 1}{r^{L-1}}$.

Using the recursive bound in Lemma 2.4,

$$D_{\mathbf{s}_1} \leq D_{(2,2)} + D_{(2,4,5,4)} + D_{(4,6,4,1)} - 2 - 4.$$

Proposition 2.5 says $(2,4,5,4)$ is a filling architecture with dimension 20, and with the dimension of the ambient space being 35 we have $D_{(4,6,4,1)} \leq 35$. It follows $D_{\mathbf{s}_1} \leq 4 + 20 + 35 - 2 - 4 = 53$.

Similarly, the recursive bound gives these inequalities:

$$\begin{aligned}
D_{\mathbf{s}_2} &\leq D_{(2,3,3)} + D_{(3,5,4)} + D_{(4,6,4,1)} - 7 \\
D_{\mathbf{s}_3} &\leq D_{(2,3,4,4)} + D_{(4,4)} + D_{(4,6,4,1)} - 8 \\
D_{\mathbf{s}_4} &\leq D_{(2,3,4,5,3)} + D_{(3,6,4,1)} - 3 \\
D_{\mathbf{s}_5} &\leq D_{(2,3,4,5,4)} + D_{(4,5)} + D_{(5,4,1)} - 9 \\
D_{\mathbf{s}_6} &\leq D_{(2,3,4,5,4)} + D_{(4,6,3)} + D_{(3,1)} - 7.
\end{aligned}$$

Using Proposition 2.5 and $D_{(m,n)} = mn$, this simplifies to:

$$\begin{aligned}
D_{\mathbf{s}_2} &\leq 9 + 24 + 35 - 7 && \leq 61 \\
D_{\mathbf{s}_3} &\leq 20 + 16 + 35 - 8 && \leq 63 \\
D_{\mathbf{s}_4} &\leq 27 + 15 - 3 && \leq 39 \\
D_{\mathbf{s}_5} &\leq 36 + 20 + 15 - 9 && \leq 62 \\
D_{\mathbf{s}_6} &\leq 36 + 30 + 3 - 7 && \leq 62.
\end{aligned}$$

Each of these dimensions is strictly less than the ambient dimension 65, hence none are filling. □

3. Search Algorithm

Our algorithm for finding MFAs exploits Lemma 2.3 and the definition of minimality. The pseudocode is provided in Algorithm 1.

The difficulty in finding MFAs is due to the size of the search space.

To efficiently find MFAs $(d_0, \mathbf{a}, d_L)$, we employ a frontier search over $\mathbf{a} = (d_1, \ldots, d_{L-1}) \in [m,n]^{L-1}$. We first prune the search space using naive dimension bounds, e.g.,

Algorithm 1 Frontier search for hidden layers' widths

Input: depth L, fixed endpoints d_0, d_L, widths range $[m, n]$, exponent r, integer B, seed seed
Initialize: $\mathcal{F}_{\min} \leftarrow \emptyset$ (minimal filling antichain)
Initialize: $\mathcal{N}_{\max} \leftarrow \emptyset$ (maximal non-filling antichain)
Initialize: rng $\leftarrow$ seed
for $t = 1$ **to** B **do**
  Propose a candidate $\mathbf{a} \in \{m, \ldots, n\}^{L-1}$ using rng
  if $\exists\, \mathbf{f} \in \mathcal{F}_{\min}$ with $\mathbf{a} \geq \mathbf{f}$ **then**
    continue (already implied filling)
  end if
  if $\exists\, \mathbf{q} \in \mathcal{N}_{\max}$ with $\mathbf{a} \leq \mathbf{q}$ **then**
    continue (already implied non-filling)
  end if
  if $(d_0, \mathbf{a}, d_L)$ is filling **then**
    $\mathcal{F}_{\min} \leftarrow \mathcal{F}_{\min} \cup \{\mathbf{a}\}$ (keep minimal)
    Remove any $\mathbf{f} \in \mathcal{F}_{\min}$ with $\mathbf{f} > \mathbf{a}$
  else
    $\mathcal{N}_{\max} \leftarrow \mathcal{N}_{\max} \cup \{\mathbf{a}\}$ (keep maximal)
    Remove any $\mathbf{q} \in \mathcal{N}_{\max}$ with $\mathbf{q} < \mathbf{a}$
  end if
end for
Output: $\mathcal{F}_{\min}$

$\dim(\mathcal{V}_{\mathbf{d},r}) \leq \sum_{i=1}^{L} d_i d_{i-1} - \sum_{i=1}^{L-1} d_i$. If there were any previous searches already performed, then we prune the search space using the results. Then, we repeatedly test $(d_0, \mathbf{a}, d_L)$ from among the remaining possibilities for filling architectures $(d_0, \mathbf{a}, d_L)$:

- If $(d_0, \mathbf{a}, d_L)$ is filling, we prune all $\mathbf{a}' > \mathbf{a}$.
- If $(d_0, \mathbf{a}, d_L)$ is not filling, we prune all $\mathbf{a}' < \mathbf{a}$.

Eventually this terminates to a finite set of candidates. We verify each candidate by computing the dimension of its subarchitectures. If verification suggests a candidate is not minimal, then we adjust our m and n and run the search again. Utilizing previous search history to inform subsequent searches allowed us to quickly prune the search space.

Remark 3.1. To determine if $(d_0, \mathbf{a}, d_L)$ is filling, our implementation of Algorithm 1, relies on finite field computations and generic choices of parameters. Accordingly, the output is correct for almost all primes and parameter choices. We then certify the result a posteriori by proving the relevant dimension bounds, as shown in Section 2.

4. Discussion

4.1. Minimal Filling Architectures

Fix $d_0 = 2$, $d_L = 1$, and $r = 2$. Our search found 13 MFAs of depth $L = 7$ with only one among them being nonunimodal: the architecture discussed in Section 2. For depth $L = 8$, we found 82 MFAs and among them, 20 were nonunimodal [2].

Table 2. MFAs for $d_0 = 2$, $d_L = 1$, $r = 2$, and $L \leq 7$.

L	MINIMAL FILLING ARCHITECTURE
$L = 2$	$(2, 2, 1)$
$L = 3$	$(2, 2, 2, 1)$
$L = 4$	$(2, 3, 3, 2, 1)$
$L = 5$	$(2, 3, 3, 3, 2, 1)$
$L = 6$	$(2, 3, 3, 4, 4, 2, 1)$
$L = 7$	$(2, 3, 4, 5, 4, 6, 4, 1), (2, 3, 4, 5, 6, 4, 2, 1)$
	$(2, 3, 3, 4, 5, 6, 3, 1), (2, 3, 4, 5, 5, 5, 2, 1)$
	$(2, 3, 3, 5, 6, 4, 4, 1), (2, 3, 3, 5, 7, 4, 2, 1)$
	$(2, 3, 3, 4, 6, 5, 2, 1), (2, 3, 4, 5, 5, 4, 4, 1)$
	$(2, 3, 4, 4, 5, 5, 4, 1), (2, 3, 3, 6, 6, 4, 3, 1)$
	$(2, 3, 3, 5, 5, 5, 4, 1), (2, 3, 4, 6, 5, 4, 3, 1)$
	$(2, 3, 5, 5, 5, 4, 3, 1)$

Fix d_0, d_L, r, and L. By Dickson's Lemma, one knows there are finitely many MFAs. Finite field computations over $p = 2^{31} - 1$ strongly suggest that every entry in Table 2 is minimal filling. A proof would follow the same strategy as Lemma 2.6, but with substantially more cases.

A natural next question is whether Table 2 is exhaustive for $L \leq 7$. Appendix B proves that this is indeed the case. For larger values of L, however, the problem is substantially more difficult. At present, there is no general method for proving completeness of such lists, since we do not know, a priori, how large the hidden widths must be in order to guarantee that all minimal filling architectures have been found.

4.2. Defectivity: losing expected expressive power

In studying the counterexample, we observed that its subarchitectures had large defect, a notion we recall next. Defectivity (for fixed activation degree r) is a measure of an architecture's loss in expressivity which cannot be attributed to expected symmetries. Presently, there does not appear to be a systematic understanding of which architectures are defective. This motivates the definition.

Definition 4.1. If $\mathbf{d} = (d_0, \ldots, d_L)$, then the *expected dimension* of the neurovariety $\mathcal{V}_{\mathbf{d},r}$, denoted $\operatorname{expdim}(\mathcal{V}_{\mathbf{d},r})$, equals the minimum of the dimension of the ambient space and $\sum_{i=1}^{L} d_i d_{i-1} - \sum_{i=1}^{L-1} d_i$. The *defect* of the neurovariety is defined to be the difference $\operatorname{expdim}(\mathcal{V}_{\mathbf{d},r}) - \dim(\mathcal{V}_{\mathbf{d},r})$.

Example 4.2 (A counterexample with $d_L = 2$)**.** Using Algorithm 1, we found a counterexample to Conjecture 1.7 where $d_0 = 2$, $d_L = 2$, $r = 2$, and $L = 9$. The PNN with architecture $\mathbf{d} = (2, 3, 4, 4, 10, 17, 11, 12, 4, 2)$ is minimal filling, and is shown in Figure 2. This can be proved using the same techniques as in our first example. For $i = 1, \ldots, L-1$, let $\mathbf{s}_i$ denote the architecture obtained

[2] The repository https://github.com/daokevin06/MFA_PNNs includes our search and implementation.

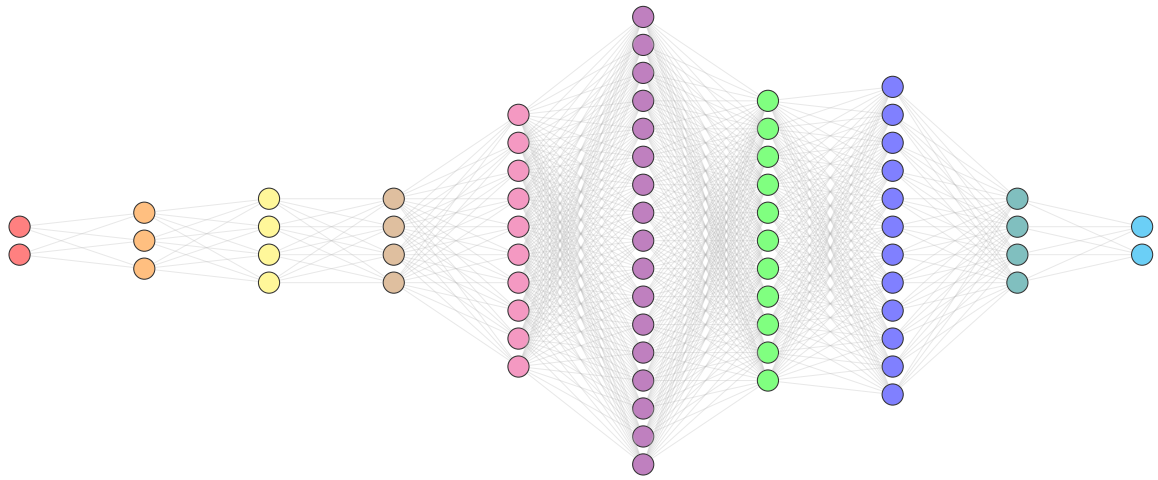

Figure 2. A counterexample to the unimodal minimal filling architecture conjecture with $d_L = 2$ ($r = 2$)

by decreasing the ith hidden layer of $\mathbf{d}$ by one. Then

$$(\mathrm{codim}(\mathcal{V}_{\mathbf{s}_i,r}))_{i=1}^{L-1} = (254, 17, 176, 5, 17, 34, 11, 124).$$

Moreover, $\mathrm{expdim}(\mathcal{V}_{\mathbf{s}_i,r}) = \dim(\mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0})^{\oplus d_L})$ holds for each i, hence $\mathrm{defect}(\mathcal{V}_{\mathbf{s}_i,r}) = \mathrm{codim}(\mathcal{V}_{\mathbf{s}_i,r})$.

Example 4.3 (A counterexample with $r = 3$)**.** A heuristic derived from the main results of (Finkel et al., 2025) is that as r increases, one expects defective architectures to become "rarer". Using Algorithm 1, we nevertheless found a counterexample to Conjecture 1.7 where $r = 3$. The architecture $\mathbf{d} = (2, 3, 7, 6, 15, 8, 1)$ is nonunimodal and minimal filling. We successfully verified it is a minimal filling architecture using the same techniques as above. The ambient dimension is 244, and the reported neurovariety dimensions are

$$(\dim(\mathcal{V}_{\mathbf{s}_i,r}))_{i=1}^{L-1} = (84, 240, 223, 233, 231).$$

So the reported defects are

$$(\mathrm{defect}(\mathcal{V}_{\mathbf{s}_i,r}))_{i=1}^{L-1} = (156, 0, 4, 2, 2).$$

Unlike Example 4.2, not all of the non-filling architectures $\mathbf{s}_i$ are defective: for some i, the expected dimension is already strictly smaller than the ambient dimension. This example shows that maximal subarchitectures of MFAs need not have large defect or have expected dimension equal to that of the ambient space.

The defects of some architectures in Table 1 turn out to be quite large. These appear to be among the first nontrivial examples of architectures exhibiting large defects. Of course, one can manufacture arbitrarily large defect by making one of the hidden layers have width one while allowing the remaining layers to be very wide. Example 4.3 shows that maximal subarchitectures of an MFA need not be defective, and even when they are defective, their defect need not be large. This motivates the study of **maximal non-filling architectures** and their defect.

It is well known that polynomial activations present a trade-off: low-degree polynomials can limit expressivity, whereas higher-degree polynomials often lead to numerical instability (Zhou et al., 2019) and gradient explosion (Goyal et al., 2020). Efforts to address these issues are already present in the literature (Hossain & Rahman, 2025). One natural response is to pair low-degree polynomial activations with architectures that are *a priori* highly expressive; in this context, understanding MFAs for power activations may inform architecture selection.

4.3. Next steps and new directions

Our results concern a classical feedforward PNN model with power activation. An important generalization is to allow broader polynomial activation families, for example activations expressed in orthogonal polynomial bases such as Chebyshev polynomials. A complementary direction is to investigate the polynomial networks (Π-nets) in (Chrysos et al., 2022) using tools from neuroalgebraic geometry, to obtain a geometric characterization of their expressivity.

Another direction is to compare training dynamics on MFAs versus maximal non-filling architectures. Several works have already investigated optimization landscapes and function-space geometry for convolutional neural networks (Kohn et al., 2022) and PNNs (Arjevani et al., 2026) through a neuroalgebraic geometry lens. But a direct comparison between MFA and maximal non-filling architectures is yet to be studied.

A new direction is to systematically characterize the boundary between nonfilling and filling architectures. This means systematically understanding minimal filling architectures. An immediate first step is to find more MFAs when $r > 2$.

Acknowledgments

We thank Kathlén Kohn and Joe Kileel for their comments on this project. We also thank the anonymous reviewers for the feedback that led to improvements in this article.

KD acknowledges that the material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2137424. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

JIR's research is partially supported by the Alfred P. Sloan Foundation and by the National Science Foundation Grant No. 2510307.

Impact Statement

This paper presents worked related to the field of Neuroalgebraic Geometry. There are potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Abo, H. On non-defectivity of certain segre-veronese varieties. *Journal of Symbolic Computation*, 2010.

Abo, H. and Brambilla, M. C. Secant Varieties of Segre-Veronese Varieties $\mathbb{P}^m \times \mathbb{P}^n$ Embedded by $\mathcal{O}(1,2)$. *Experimental Mathematics*, 2009.

Alexander, J. and Hirschowitz, A. Polynomial interpolation in several variables. *J. Algebraic Geom.*, 1995. ISSN 1056-3911,1534-7486.

Arjevani, Y., Bruna, J., Kileel, J., Polak, E., and Trager, M. Geometry and Optimization of Shallow Polynomial Networks. *SIAM J. Appl. Algebra Geom.*, 10(2):174–209, 2026. ISSN 2470-6566.

Brandenburg, M.-C., Loho, G., and Montufar, G. The real tropical geometry of neural networks for binary classification. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Breiding, P., Kohn, K., and Sturmfels, B. *Metric algebraic geometry*, volume 53 of *Oberwolfach Seminars*. Birkhäuser/Springer, Cham, 2024. ISBN 978-3-031-51461-6; 978-3-031-51462-3.

Chrysos, G. G., Moschoglou, S., Bouritsas, G., Deng, J., Panagakis, Y., and Zafeiriou, S. Deep polynomial neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8): 4021–4034, 2022.

Finkel, B., Rodriguez, J. I., Wu, C., and Yahl, T. Activation degree thresholds and expressiveness of polynomial neural networks. *Algebraic Statistics*, 16:113–130, 2025.

Goyal, M., Goyal, R., and Lall, B. Improved Polynomial Neural Networks with Normalised Activations. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2020.

Henry, N., Marchetti, G. L., and Kohn, K. Geometry of lightning self-attention: Identifiability and dimension. In *International Conference on Learning Representations*, 2025.

Hossain, F. A. and Rahman, T. A Training Framework for Optimal and Stable Training of Polynomial Neural Networks. *arXiv:2505.11589*, 2025.

Kileel, J., Trager, M., and Bruna, J. On the Expressive Power of Deep Polynomial Neural Networks. In *Advances in Neural Information Processing Systems*, 2019.

Kohn, K., Merkh, T., Montúfar, G., and Trager, M. Geometry of linear convolutional networks. *SIAM J. Appl. Algebra Geom.*, 6(3):368–406, 2022. ISSN 2470-6566.

Kubjas, K., Li, J., and Wiesmann, M. Geometry of Polynomial Neural Networks. *Algebraic Statistics*, 15(2): 295–328, 2024.

Marchetti, G. L., Shahverdi, V., Mereta, S., Trager, M., and Kohn, K. Position: Algebra Unveils Deep Learning – An Invitation to Neuroalgebraic Geometry. In *Proceedings of the 42nd International Conference on Machine Learning*. PMLR, 2025.

Massarenti, A. and Mella, M. The Alexander-Hirschowitz theorem for neurovarieties. *arXiv:2511.19703*, 2025.

Michałek, M. and Sturmfels, B. *Invitation to nonlinear algebra*, volume 211 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2021. ISBN 978-1-4704-5367-1.

Usevich, K., Borsoi, R., Dérand, C., and Clausel, M. Identifiability of deep polynomial neural networks. In *Advances in Neural Information Processing Systems*, volume 38, pp. 81809–81858. Curran Associates, Inc., 2025.

Zhang, L., Naitzat, G., and Lim, L.-H. Tropical geometry of deep neural networks. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5824–5832. PMLR, 10–15 Jul 2018.

Zhou, J., Qian, H., Lu, X., Duan, Z., Huang, H., and Shao, Z. Polynomial activation neural networks: Modeling, stability analysis and coverage BP-training. *Neurocomputing*, 359:227–240, 2019. ISSN 0925-2312.

A. Jacobians, Dimension, and Finite Fields

In this appendix, we provide some background for Lemma 2.2 and finite field computations.

Definition A.1. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a polynomial map i.e. $F(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$ where the f_i are polynomials in n-variables. The **Jacobian** of F is defined to be the $m \times n$-matrix

$$\operatorname{Jac} F = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

This is an $m \times n$-matrix whose entries are polynomials in n-variables. The **rank of** $\operatorname{Jac} F$ **at a point** $x \in \mathbb{R}^n$ is the rank of the $m \times n$-matrix with real entries $\operatorname{Jac} F(x)$ obtained by evaluating each polynomial entry of $\operatorname{Jac} F$ at the point $x \in \mathbb{R}^n$.

In algebraic geometry, a property that holds outside a proper closed subvariety is said to hold *generically*. The key fact regarding $\operatorname{Jac} F(x)$ is that there exists a Zariski closed subvariety $Z \subseteq \mathbb{R}^n$ such that the rank of $\operatorname{Jac} F(x)$ is constant for $x \in \mathbb{R}^n \setminus Z$.

Theorem A.2. *If $F : \mathbb{R}^n \to \mathbb{R}^m$ is a polynomial map, then there exists a Zariski open subset $U \subseteq \mathbb{R}^n$ so that for all $x \in U$,*

$$\operatorname{rank}(\operatorname{Jac} F(x)) = \dim \overline{\operatorname{Im}(F)}.$$

where the bar denotes taking the Zariski closure inside $\mathbb{R}^m$. Furthermore, outside of U, the rank can only drop.

Proof. See Lemma A.1 of (Usevich et al., 2025). □

The value of $\operatorname{rank}(\operatorname{Jac} F(x))$ on U in the previous theorem is referred to as the **generic rank**. The previous theorem implies, for generic $x \in \mathbb{R}^n$,

$$\operatorname{rank}(\operatorname{Jac} F(x)) = \dim \overline{\operatorname{Im}(F)}.$$

Accordingly, if $\operatorname{Jac} \Psi_{\mathbf{d},r}(\theta)$ has full row rank for some θ, then $\mathcal{V}_{\mathbf{d},r}$ is filling. In practice, such a θ is found by a random choice of parameters. For more details, the reader can refer to Section 3.3 of (Kileel et al., 2019).

Example A.3 (Linear maps)**.** Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map defined by $F(x) = Ax$ for some $m \times n$ matrix $A \in \mathbb{R}^{m \times n}$. Then

$$\operatorname{Jac} F(x) = A$$

for every $x \in \mathbb{R}^n$, so the rank of the Jacobian is constant and equal to $\operatorname{rank}(A)$. Since the image of F is the linear subspace spanned by the columns of A, we also have

$$\dim \overline{\operatorname{Im}(F)} = \dim \operatorname{Im}(A) = \operatorname{rank}(A).$$

Thus the theorem holds in this case.

Example A.4 (A nonlinear map)**.** Consider the map $F : \mathbb{R} \to \mathbb{R}$, $F(x) = x^2$. Here

$$\operatorname{Jac} F(x) = [2x].$$

Hence

$$\operatorname{rank}(\operatorname{Jac} F(x)) = \begin{cases} 1, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Thus the generic rank is 1. On the other hand,

$$\operatorname{Im}(F) = [0, \infty),$$

whose Zariski closure in $\mathbb{R}$ is all of $\mathbb{R}$. Therefore

$$\dim \overline{\operatorname{Im}(F)} = 1,$$

in agreement with the theorem.

Example A.5 (A quadratic map with one-dimensional image)**.** Let $q(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$. Consider the polynomial map $F : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$F(x_1, x_2, x_3) = (q(x_1, x_2, x_3), 2q(x_1, x_2, x_3).$$

Then

$$\operatorname{Jac} F(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 \\ 4x_1 & 4x_2 & 4x_3 \end{pmatrix},$$

whose second row is always twice the first. Therefore the Jacobian has generic rank 1. The image of F lies in the line

$$\{(u, 2u) : u \in \mathbb{R}\} \subseteq \mathbb{R}^2,$$

and its Zariski closure is exactly this line, which has dimension 1.

Corollary A.6. *If $F : \mathbb{R}^n \to \mathbb{R}^m$ is a polynomial map with **integer coefficients**, then there exists a Zariski dense open subset $U \subseteq \mathbb{R}^n$ such that for every $x \in \mathbb{Z}^n \cap U$, one has*

$$\operatorname{rank}(\operatorname{Jac} F(x)) = \operatorname{rank}(\operatorname{Jac} F(x) \bmod p)$$

for all but finitely many primes p.

In Corollary A.6, $\operatorname{Jac} F(x) \bmod p$ is obtained by evaluating the entries of the matrix $\operatorname{Jac} F$ at x and then reducing modulo p to a $\mathbb{F}_p$-valued matrix. Since reduction modulo p can only ever reduce the rank, if $x \in \mathbb{Z}^n$ and p a prime, then $\operatorname{rank}(\operatorname{Jac} F(x)) \geq \operatorname{rank}(\operatorname{Jac} F(x) \bmod p)$.

The parameter map for neurovarieties always has integer coefficients. So, if one can find a prime p and a parameter θ in the parameter space so that $\operatorname{Jac}(\Psi_{\mathbf{d},r}(\theta) \bmod p)$ rank equal to the dimension of the ambient space, then $\mathcal{V}_{\mathbf{d},r}$ is filling. Furthermore, if one can exhibit an upper bound $\dim \mathcal{V}_{\mathbf{d},r} \leq B$ and then find a prime p and parameter $\theta \in \mathbb{Z}^n$ for which $\operatorname{rank} \operatorname{Jac}(\Psi_{\mathbf{d},r}(\theta) \bmod p) = B$, one proves $\dim \mathcal{V}_{\mathbf{d},r} = B$.

Example A.7 (Linear maps revisited)**.** Suppose $F(x) = Ax$ with $A \in \mathbb{Z}^{m \times n}$. Then

$$\operatorname{Jac} F(x) = A$$

for all $x \in \mathbb{Z}^n$. Let $r = \operatorname{rank}(A)$. The rank of $A \bmod p$ drops below r exactly when all $r \times r$ minors of A vanish modulo p. Hence any such prime p must divide every nonzero $r \times r$ minor of A. Since there are only finitely many minors, and each nonzero minor has only finitely many prime divisors, there are only finitely many primes for which the rank drops. Thus

$$\operatorname{rank}(A) = \operatorname{rank}(A \bmod p)$$

for all but finitely many primes p.

Example A.8. For the architecture $\mathbf{d} = (2, 1, 1)$ and $r = 2$, the parameter map is given by

$$\Psi_{\mathbf{d},r} : \mathbb{R}^{1\times 2} \times \mathbb{R}^{1\times 1} \to \mathbb{R}^3 \qquad \left(\begin{bmatrix} w_{11} & w_{12} \end{bmatrix}, \begin{bmatrix} w_2 \end{bmatrix}\right) \mapsto (w_2 w_{11}^2, 2w_2 w_{11} w_{12}, w_2 w_{12}^2).$$

Its Jacobian matrix is

$$\operatorname{Jac}(\Psi_{\mathbf{d},r}) = \begin{bmatrix} 2w_2 w_{11} & 0 & w_{11}^2 \\ 2w_2 w_{12} & 2w_2 w_{11} & 2w_{11} w_{12} \\ 0 & 2w_2 w_{12} & w_{12}^2 \end{bmatrix}.$$

The determinant of this matrix is identically zero, so $\operatorname{Jac}(\Psi_{\mathbf{d},r})(w_{11}, w_{12}, w_2)$ has rank at most 2 for all $(w_{11}, w_{12}, w_2) \in \mathbb{R}^3$. Moreover, at the point $(w_{11}, w_{12}, w_2) = (1, 0, 1)$, the Jacobian is

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

which has rank 2. Hence the generic rank is 2, and therefore

$$\dim \mathcal{V}_{\mathbf{d},r} = 2.$$

The rank of the Jacobian at a point drops below 2 exactly when all nine 2×2 minors vanish. For the integer point $(w_{11}, w_{12}, w_2) = (1, 0, 1)$, the reduced Jacobian modulo p has rank 2 for every prime $p \neq 2$. Thus, a single computation modulo any odd prime certifies that the generic rank is 2.

Lastly, the task of efficiently computing $\operatorname{Jac}\Psi_{\mathbf{d},r}$ is resolved by backpropagation. The interested reader can refer to Section 6.4 of (Kubjas et al., 2024) or to Section 3.3 of (Kileel et al., 2019). More broadly, for textbook references on applications of algebraic geometry we refer to (Michałek & Sturmfels, 2021; Breiding et al., 2024).

B. Verifying Completeness of List of MFAs for $L \leq 7$

In this appendix, we prove Table 2 is a complete list of MFAs for $d_0 = 2, d_L = 1$, and $L \leq 7$.

Lemma B.1 (Dickson's Lemma)**.** *Let $\mathbb{N}$ be the set of positive integers. If $(\mathbb{N}^r, \leq)$ is the partially ordered set where $(a_1, \ldots, a_r) \leq (b_1, \ldots, b_r)$ if and only if $a_i \leq b_i$ for all i, then any subset $\mathcal{S} \subseteq \mathbb{N}^r$ has at most finitely many minimal elements.*

Dickson's lemma implies that the set of minimal filling architectures is finite for fixed r, L, d_0, d_L. Verifying that the list for $L \leq 6$ is exhaustive is not difficult. For each $L \leq 6$, there is a unique MFA and the regions consisting of incomparable architectures are never filling by repeated application of the recursive bound.

Recall that we use $D_{\mathbf{d}}$ to denote $\dim \mathcal{V}_{\mathbf{d},r}$ and that we are fixing $d_0 = 2, d_L = 1$.

Lemma B.2. *Table 2 is complete for $L \leq 6$.*

Proof. Since the arguments for $L \leq 5$ are analogous and shorter, we present only the case $L = 6$. Using the dimension of the ambient space and the recursive bound, we can impose conditions on the hidden layers' widths d_i which must be satisfied for the architecture to be filling.

We claim that if $\mathbf{d} = (2, d_1, d_2, d_3, d_4, d_5, 1)$ is filling, then $d_1 \geq 3,\ d_2 \geq 3,\ d_3 \geq 4\ , d_4 \geq 4,$ and $d_5 \geq 2$. It is equivalent to show that if one of these inequalities fail, then $\mathbf{d}$ is not filling. The proof of these inequalities are quite similar and so we only present the case $d_4 \geq 4$ fails. We must show that if $d_4 < 4$, then $\mathbf{d}$ cannot be filling. Indeed, if $d_4 = 3$, then $D_{\mathbf{d}} \leq D_{(2,d_1,d_2,d_3,3)} + D_{(3,d_5,1)} - 3 \leq 27 + 6 - 3 = 30$. The ambient space $\operatorname{Sym}_{2^{L-1}}(\mathbb{R}^2)$ has dimension 33. So, $\mathbf{d}$ cannot be a filling architecture if $d_4 < 4$.

These inequalities imply that if $(2, \mathbf{a}, 1)$ is a filling architecture, then $\mathbf{a} \in \mathcal{S}$ where $\mathcal{S} = [3, \infty) \times [3, \infty) \times [4, \infty) \times [4, \infty) \times [2, \infty) \subseteq \mathbb{N}^5$. The minimal element in $\mathcal{S}$ with respect to the partial ordering $(\mathbb{N}^5, \leq)$ given by coordinate-wise comparison is $\mathbf{a} = (3, 3, 4, 4, 2)$. We observed by computer verification that $(2, 3, 3, 4, 4, 2, 1)$ is filling. Thus, it must be minimal filling and it is the only MFA for $L = 6$. □

Lemma B.3. *Assume that*

$$D_{(5,5,3,1)} = 32, \qquad D_{(3,5,6,4,3,1)} = 58, \qquad D_{(4,5,5,3,1)} = 47, \qquad D_{(5,4,3,1)} = 25.$$

Then Table 2 is complete for $L = 7$.

Proof. If we take our list of thirteen MFAs, then we can determine the regions which have not yet been checked for the presence of a MFA. There are fifteen such infinite rectangular regions:

Region 1: $[3, \infty) \times [3, 3] \times [4, \infty) \times [4, 4] \times [4, \infty) \times [2, \infty)$
Region 2: $[3, \infty) \times [3, 3] \times [4, 4] \times [4, 5] \times [4, 5] \times [2, \infty)$
Region 3: $[3, \infty) \times [3, 3] \times [4, \infty) \times [4, 5] \times [4, 4] \times [2, \infty)$
Region 4: $[3, \infty) \times [3, 3] \times [4, \infty) \times [4, 5] \times [4, 5] \times [2, 3]$
Region 5: $[3, \infty) \times [3, 3] \times [4, 5] \times [4, 6] \times [4, 4] \times [2, 3]$
Region 6: $[3, \infty) \times [3, 3] \times [4, \infty) \times [4, 6] \times [4, 4] \times [2, 2]$
Region 7: $[3, \infty) \times [3, 3] \times [4, \infty) \times [4, 5] \times [4, \infty) \times [2, 2]$
Region 8: $[3, \infty) \times [3, \infty) \times [4, 4] \times [4, 4] \times [4, \infty) \times [2, \infty)$
Region 9: $[3, \infty) \times [3, \infty) \times [4, 4] \times [4, 5] \times [4, 5] \times [2, 3]$
Region 10: $[3, \infty) \times [3, \infty) \times [4, 4] \times [4, \infty) \times [4, 4] \times [2, \infty)$
Region 11: $[3, \infty) \times [3, \infty) \times [4, 4] \times [4, 5] \times [4, \infty) \times [2, 2]$
Region 12: $[3, \infty) \times [3, \infty) \times [4, \infty) \times [4, 4] \times [4, 5] \times [2, \infty)$
Region 13: $[3, \infty) \times [3, 4] \times [4, 5] \times [4, 5] \times [4, 4] \times [2, 3]$
Region 14: $[3, \infty) \times [3, \infty) \times [4, \infty) \times [4, 5] \times [4, 4] \times [2, 2]$
Region 15: $[3, \infty) \times [3, \infty) \times [4, \infty) \times [4, 4] \times [4, \infty) \times [2, 3]$

It suffices to show that no architecture $(2, \mathbf{a}, 1)$ with $\mathbf{a}$ lying in one of these regions is filling. We use Lemma 2.4 and expected dimension to derive upper bounds on the dimensions of $\mathcal{V}_{(2,\mathbf{a},1)}$. This can be done systematically. For example, for Region 1, one considers $D_{(2,d_1,3,d_3,4,d_5,d_6,1)}$, and using the recursive bound and ambient dimension,

$$D_{(2,d_1,3)} + D_{(3,d_3,4)} + D_{(4,d_5,d_6,1)} - 7 \leq 9 + 24 + 35 - 7 = 61 < 65.$$

We found that similar bounds can be proven for every region except Regions 4, 5, 9, and 13. For Regions 4, 5, 9, 13, better estimates on the dimensions are needed and this need arises from influence of defective architectures e.g. $(5, 5, 3, 1)$ has defect 3. For Regions 4, 5, 9, 13, we respectively have

$$\begin{aligned}
D_{(2,d_1,3,d_3,5,5,3,1)} &\leq & D_{(2,d_1,3)} + D_{(3,d_3,5)} + D_{(5,5,3,1)} - 3 - 5 &\leq & 9 + 30 + 32 - 3 - 5 = 63 \\
D_{(2,d_1,3,5,6,4,3,1)} &\leq & D_{(2,d_1,3)} + D_{(3,5,6,4,3,1)} - 3 &\leq & 9 + 58 - 3 = 64 \\
D_{(2,d_1,d_2,4,5,5,3,1)} &\leq & D_{(2,d_1,d_2,4)} + D_{(4,5,5,3,1)} - 4 &\leq & 20 + 47 - 4 = 63 \\
D_{(2,d_1,4,5,5,4,3,1)} &\leq & D_{(2,d_1,4,5)} + D_{(5,5)} + D_{(5,4,3,1)} - 10 &\leq & 24 + 25 + 25 - 10 = 64.
\end{aligned}$$

We also use $D_{(2,d_1,4,5)} \leq 24$ for $d_1 \geq 3$ which follows from the main results of (Massarenti & Mella, 2025). □

To complete the verification that our list of MFAs for $L = 7$ is complete, we must verify the dimension counts assumed in Lemma B.3. To do so, we will use the Alexander-Hirschowitz Theorem and some results on nondefectivity of secant varieties of Segre-Veronese varieties.

Theorem B.4. *(Alexander & Hirschowitz, 1995) If $\mathbf{d} = (d_0, d_1, 1)$, the dimension of $\mathcal{V}_{\mathbf{d},r}$ is given by $\min\left(d_0 d_1, \binom{d_0+r-1}{r}\right)$, except for the following cases:*

- $r = 2, 2 \leq d_1 \leq d_0 - 1$, *and* $\dim \mathcal{V}_{\mathbf{d},r} = d_0 d_1 - \binom{d_1}{2}$,
- $r = 3, d_0 = 5, d_1 = 7$, *and* $\dim \mathcal{V}_{\mathbf{d},r} = 34$
- $r = 4, d_0 = 3, d_1 = 5$, *and* $\dim \mathcal{V}_{\mathbf{d},r} = 14$,
- $r = 4, d_0 = 4, d_1 = 9$, *and* $\dim \mathcal{V}_{\mathbf{d},r} = 35$
- $r = 4, d_0 = 5, d_1 = 15$, *and* $\dim \mathcal{V}_{\mathbf{d},r} = 74$.

Theorem B.5. *(Abo, 2010) Let $r = 2$. If $\mathbf{d} = (d_0, d_1, d_2)$ satisfies either $d_2 = d_0+1$ or $d_2 = d_0$, and $(d_0, d_1, d_2) \neq (4, 6, 5)$, then $\mathcal{V}_{\mathbf{d},r}$ is not defective.*

Theorem B.6. *If $r = 2$ and $\mathbf{d} = (d_0, d_1, d_2)$ satisfies $\binom{d_0-1+r}{r} - d_0 < d_1 < \min\left\{d_2, \binom{d_0-1+r}{r}\right\}$, then $\mathcal{V}_{\mathbf{d},r}$ is defective. In particular, this applies to architecture $(3, 5, 6)$ with $r = 2$.*

Proof. See Remark 4.4 of (Abo & Brambilla, 2009). □

The following lemma supplies the four dimension computations used in Lemma B.3, and hence completes the proof that the table of MFAs is complete for $L \leq 7$, $d_0 = 2$, $d_L = 1$, and $r = 2$.

Lemma B.7. *If $r = 2$, then $D_{(5,5,3,1)} = 32$, $D_{(3,5,6,4,3,1)} = 58$, $D_{(4,5,5,3,1)} = 47$, and $D_{(5,4,3,1)} = 25$.*

Proof. To show each of the following equalities, we need to find an upper bound as these are the reported dimensions obtained from the backpropagation algorithm of (Kubjas et al., 2024).

$D_{(5,5,3,1)} = 32$. Using the recursive bound and Theorem B.4 to get $D_{(5,3,1)} = 12$, one has

$$D_{(5,5,3,1)} \leq D_{(5,5)} + D_{(5,3,1)} - 5 = 25 + D_{(5,3,1)} - 5 = 20 + D_{(5,3,1)} = 32.$$

$D_{(3,5,6,4,3,1)} = 58$. Use Theorem B.4 to get $D_{(4,3,1)} = 9$. Then, utilizing the recursive bound,

$$D_{(3,5,6,4,3,1)} \leq D_{(3,5,6,4)} + D_{(4,3,1)} - 4 \leq D_{(3,5,6,4)} + D_{(4,3,1)} - 4 = D_{(3,5,6,4)} + 9 - 4 = D_{(3,5,6,4)} + 5.$$

It remains to show $D_{(3,5,6,4)} = 53$. By the recursive bound, we have $D_{(3,5,6,4)} \leq D_{(3,5,6)} + D_{(6,4)} - 6 = D_{(3,5,6)} + 18$. Theorem B.6 implies $D_{(3,5,6)} < 36$. Therefore, $D_{(3,5,6,4,3,1)} \leq 58$.

$D_{(4,5,5,3,1)} = 47$. Since Theorem B.4 and Theorem B.5 imply $D_{(5,3,1)} = 12$ and $D_{(4,5,5)} = 40$, the recursive bound provides

$$D_{(4,5,5,3,1)} \leq D_{(4,5,5)} + D_{(5,3,1)} - 5 = 40 + 12 - 5 = 47.$$

$D_{(5,4,3,1)} = 25$. Theorem B.4 says $D_{(4,3,1)} = 9$, and so the recursive bound gives

$$D_{(5,4,3,1)} \leq D_{(5,4)} + D_{(4,3,1)} - 4 = 16 + D_{(4,3,1)} = 25.$$

□

C. Parameter choices for Lemma 2.2

The weight matrices for the computation in Lemma 2.2 are provided below. We used the prime $p = 101$.

$$W_1 = \begin{bmatrix} 94 & 6 \\ 9 & 7 \\ 58 & 11 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 96 & 20 & 67 \\ 83 & 52 & 74 \\ 30 & 41 & 2 \\ 3 & 90 & 1 \end{bmatrix}$$

$$W_3 = \begin{bmatrix} 23 & 36 & 99 & 71 \\ 27 & 94 & 56 & 3 \\ 91 & 50 & 65 & 62 \\ 24 & 20 & 30 & 76 \\ 41 & 49 & 76 & 81 \end{bmatrix}$$

$$W_4 = \begin{bmatrix} 94 & 55 & 89 & 90 & 60 \\ 100 & 41 & 6 & 49 & 83 \\ 78 & 56 & 50 & 3 & 56 \\ 16 & 46 & 15 & 10 & 16 \end{bmatrix}$$

$$W_5 = \begin{bmatrix} 51 & 37 & 71 & 85 \\ 97 & 23 & 57 & 56 \\ 12 & 47 & 41 & 60 \\ 31 & 60 & 63 & 15 \\ 54 & 2 & 57 & 60 \\ 40 & 2 & 34 & 83 \end{bmatrix}$$

$$W_6 = \begin{bmatrix} 57 & 90 & 8 & 19 & 55 & 36 \\ 8 & 35 & 94 & 89 & 25 & 36 \\ 8 & 55 & 6 & 69 & 14 & 28 \\ 59 & 7 & 85 & 82 & 60 & 4 \end{bmatrix}$$

$$W_7 = \begin{bmatrix} 99 & 42 & 87 & 98 \end{bmatrix}$$