Research Highlight: The Unimodal Minimal Filling Architecture Conjecture

Kevin Dao and Jose Israel Rodriguez

A starting point for understanding neural network expressiveness is the case of *deep linear networks*. In such a network, the parameters are matrices. A function expressed by the network takes an input vector $x \in \mathbb{R}^{d_0}$ and repeatedly multiplies it by matrices W_i:

$$x \longmapsto W_L W_{L-1} \cdots W_1 x \in \mathbb{R}^{d_L}.$$

The depth of the network is L, the number of matrices. The sizes of these matrices determine the network's *architecture*. That is, if W_i is a $d_i \times d_{i-1}$ matrix, then the network has architecture $(d_0, d_1, \ldots, d_L)$. Here d_0 is the input dimension, d_L is the output dimension, and $d_1, \ldots, d_{L-1}$ are the widths of the *hidden layers*.

An architecture that can express every linear map from $\mathbb{R}^{d_0}$ to $\mathbb{R}^{d_L}$ is called *filling*. For linear networks, filling is determined by linear algebra. When the network has a single hidden layer, this is the familiar rank-factorization problem. A matrix factored through a hidden layer of width h has the form

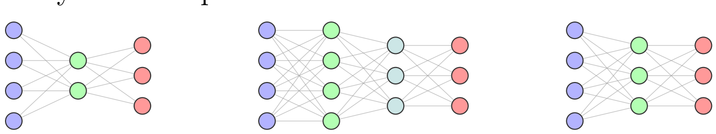
$$A = W_2 W_1, \text{ where } W_1 \in \mathbb{R}^{h \times d_0} \text{ and } W_2 \in \mathbb{R}^{d_L \times h}.$$

Such a matrix has rank at most h. Therefore, to represent every linear map from $\mathbb{R}^{d_0}$ to $\mathbb{R}^{d_L}$,

$$\text{the hidden widths must be at least } \min(d_0, d_L).$$

Moreover, a deep linear architecture is filling precisely when every hidden layer is at least this large.

Example. The neural networks below have architectures $(4, 2, 3)$, $(4, 4, 3, 3)$, $(4, 3, 3)$. They express functions from $\mathbb{R}^4$ to $\mathbb{R}^3$. The architecture on the left is not filling. The other two networks are filling; this is because every linear map from $\mathbb{R}^4$ to $\mathbb{R}^3$ has rank at most three.



Polynomial neural networks (PNNs) generalize this picture by inserting an activation function between linear maps. In the setting studied in [2] the activation is a power function applied coordinatewise, such as $z \longmapsto z^r$.

PNN Example. Fix the architecture to be $(2, d_1, 1)$, and denote the input variable by $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. If the parameters are chosen to be

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

then the resulting homogeneous polynomial function expressed by the network is of degree r:

$$x \mapsto (x_1 + x_2)^r + (x_1 + 2x_2)^r + \cdots + (x_1 + d_1 x_2)^r.$$

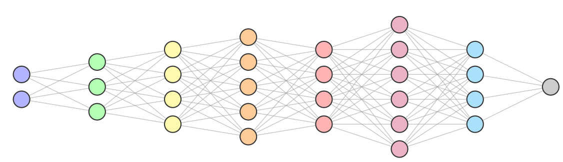
For a fixed architecture, varying the entries of the weight matrices determines a polynomial map from the network parameters to a space of homogeneous polynomial functions. The image is the *neuromanifold*; the smallest algebraic set containing it is the *neurovariety*. They agree in the linear case, and always have the same dimension, which can be computed as the maximal rank of the Jacobian matrix of the parameterization.

In the PNN setting, an architecture is *filling* if its neurovariety is the entire ambient space of homogeneous polynomial functions of the appropriate degree. It is *minimal filling* if it is filling and stops being filling as soon as any hidden-layer width is decreased.

The linear story suggests that minimal filling architectures should have a simple shape. Kileel, Trager, and Bruna formulated the *unimodal minimal filling architecture conjecture*, which predicts that, once the input and output widths are fixed, every minimal filling PNN architecture should have hidden-layer widths that increase up to a peak and then decrease.

Recent work [1] by Kevin Dao and Jose Israel Rodriguez, presented at CoLoRAI ICML 2026, shows that this natural expectation fails. The example was found through a frontier search and certified using symbolic computation and previous results on neurovarieties. Their work also established the exhaustiveness of some searches through algebraic geometry results.

The first counterexample. Using a quadratic activation function $z \mapsto z^2$, the architecture $(2, 3, 4, 5, 4, 6, 4, 1)$ is minimal filling but not unimodal.



The counterexample demonstrates that efficient expressiveness need not follow the simple rise-and-fall pattern suggested by classical intuition. It opens the door to a sharper algebraic-geometric understanding of neural network architecture design.

[1] K. Dao and J. I. Rodriguez, *Minimal filling architectures of polynomial neural networks: Counterexamples, frontier search, and defects*, in ICML'26 workshop on CoLoRAI, 2026.

[2] J. Kileel, M. Trager, and J. Bruna, *On the expressive power of deep polynomial neural networks*, Advances in neural information processing systems, 32 (2019).