Identifiability of Deep Polynomial Neural Networks

Konstantin Usevich[†], Ricardo Borsoi[†],
Clara Dérand[†], Marianne Clausel[*]

[†]CNRS, [*]University of Lorraine, CRAN, Nancy, France

4 December 2025

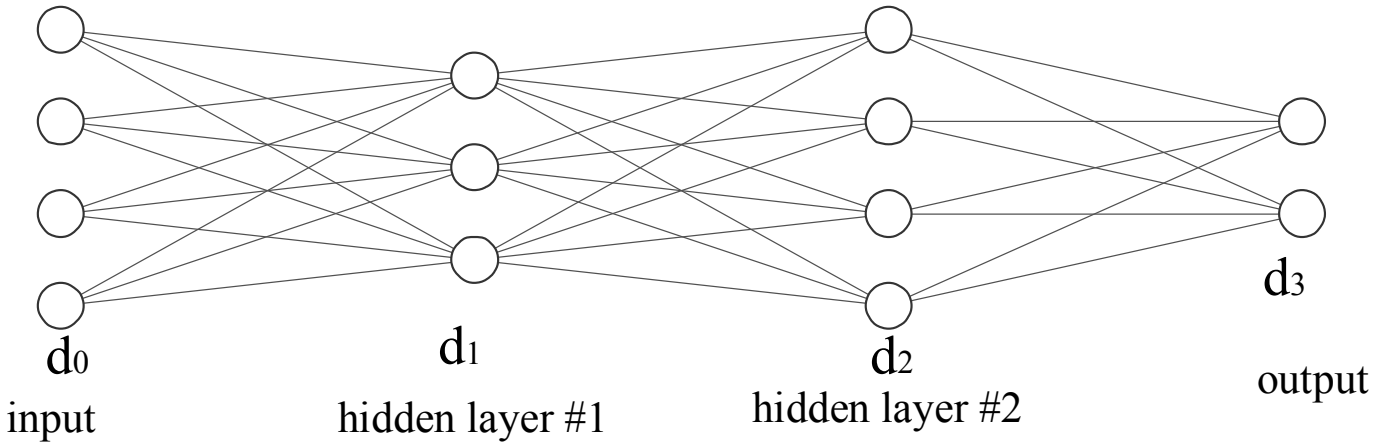
My coauthors

Ricardo Borsoi

Clara Dérand

Marianne Clausel

Feedforward neural networks: identifiability



$$f = f_L \circ \sigma_{L-1} \circ f_{L-1} \circ \sigma_{L-2} \circ \cdots \circ \sigma_1 \circ f_1 : \mathbb{R}^{d_0} \to \mathbb{R}^{d_L}$$

σ_ℓ — activation functions, $f_\ell(\mathbf{x}) = \mathbf{W}_\ell \mathbf{x} + \mathbf{b}_\ell$ — linear layers

Question: When is the NN representation (weights/biases) of f unique, except trivial symmetries (e.g., neuron permutations) ?

Identifiability of NN architecture $\stackrel{\text{def}}{=}$ uniqueness for generic $\mathbf{W}_\ell, \mathbf{b}_\ell$

Why is identifiability important?

- learning, generalization
- explainability, interpretability
- optimization landscape, dynamics
- geometry/symmetries of NNs
- merging of pretrained representations:

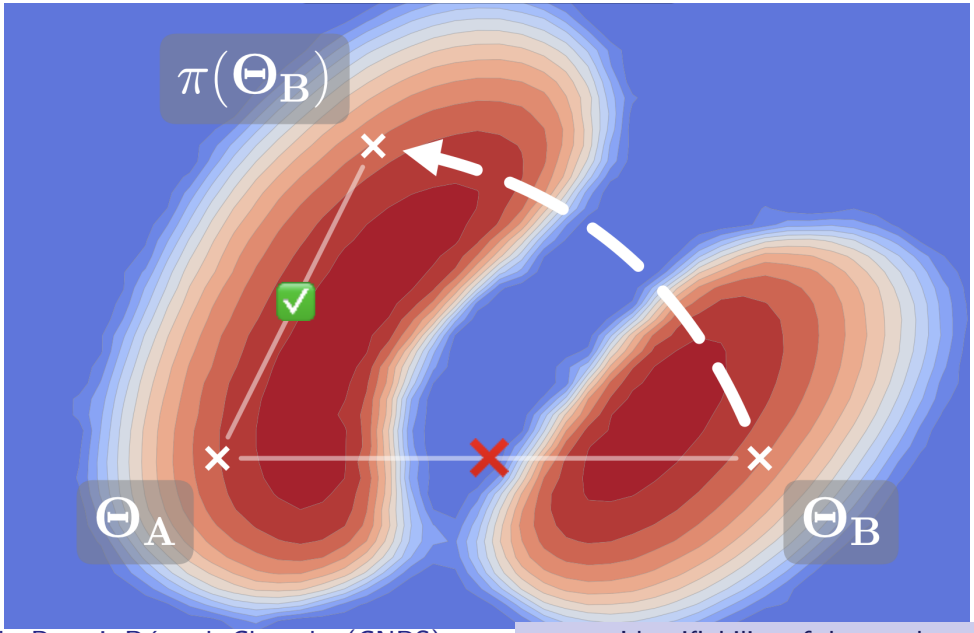


Figure: [Ainsworth et al., ICLR'23]

NeurIPS'25:

TUE 2 DEC

Tutorial:
Explain AI Models: Methods and Opportunities in Explainable AI,

⋮

Tutorial:
Foundations of Tensor/Low-Rank Computations for AI

Tutorial:
Model Merging: Theory, Practice and Applications

⋮

Tutorial:
Recent Developments in Geometric Machine Learning:

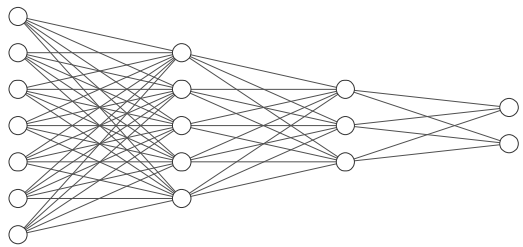
Identifiability of NNs: what is known

Many powerful results for shallow case ($L = 2$): [Sussmann, 1992], ..., [Janzamin et al. 2015], [Fornasier et al. 2021]...

Much fewer for deep networks ($L \geq 3$):

- tanh, sigmoidal: **[Fefferman, 1994]** [Vlačić, Bölcskei, 2022]
- ReLU: [Bona-Pelissier et al. 2022-23] local identifiability for

  pyramidal $d_0 \geq d_1 \geq \cdots d_L \geq 2$



- polynomial NNs: equiwidth/very high degree [Finkel et al. 2025], pyramidal case conjectured [Kubjas et al. 2024]...

This talk: shallow $\rightarrow$ deep for PNN

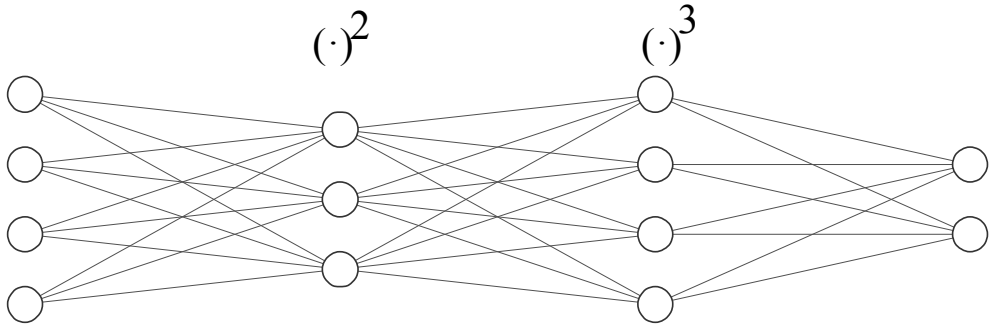
Polynomial neural nets

[Kileel et al., NeurIPS'19]:

$$f = f_L \circ \rho_{r_{L-1}} \circ f_{L-1} \circ \rho_{r_{L-2}} \circ \cdots \circ \rho_{r_1} \circ f_1,$$

- $f_\ell(\mathbf{x}) = \mathbf{W}_\ell \mathbf{x} + \mathbf{b}_\ell$, weights $\mathbf{W}_\ell \in \mathbb{R}^{d_\ell \times d_{\ell-1}}$, biases $\mathbf{b}_\ell \in \mathbb{R}^{d_\ell}$
- layer-wise monomial activation functions: $\rho_r(\mathbf{z}) := (z_1^r, \ldots, z_d^r)$
- PNN architecture:
  - widths: $\mathbf{d} = (d_0, \ldots, d_L)$
  - degrees: $\mathbf{r} = (r_1, \ldots, r_{L-1})$

Example.
$\mathbf{d} = (4, 3, 4, 2)$,
$\mathbf{r} = (2, 3)$



Why polynomials

- properties of PNNs $\leftrightarrow$ (algebraic) geometry
- shallow PNN $\leftrightarrow$ tensor decompositions
- active community of algebraic methods for NNs:

NeurIPS'19

On the Expressive Power of Deep Polynomial Neural Networks

Joe Kileel* Princeton University
Matthew Trager* New York University
Joan Bruna New York University

ICML'25

***Position:* Algebra Unveils Deep Learning**
An Invitation to Neuroalgebraic Geometry

Giovanni Luca Marchetti [*1] **Vahid Shahverdi** [*1] **Stefano Mereta** [*1] **Matthew Trager** [*2] **Kathlén Kohn** [*1]

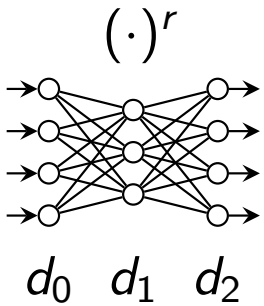
Trivial ambiguities (symmetries) of PNN

Simpler case. hPNN = polynomial neural network without bias terms

$$f = \underbrace{\mathbf{W}_L \circ \rho_{r_{L-1}} \circ \mathbf{W}_{L-1} \circ \rho_{r_{L-2}} \circ \cdots \circ \rho_{r_1} \circ \mathbf{W}_1}, \qquad \begin{array}{c} \overset{d_{\ell-1}}{} \\ d_\ell \boxed{\mathbf{W}_\ell} \end{array}$$

homogeneous polynomial (all terms of same degree)

Equivalent representations: f is invariant under
(a) neuron permutations; (b) neuron rescaling ($(az)^r = a^r z^r$)



Example.
$\mathbf{f}(\mathbf{x}) = \mathbf{W}_2 \rho_r(\mathbf{W}_1 \mathbf{x})$,

(a)	(b)
$\mathbf{W}_2' = \mathbf{W}_2 \mathbf{P}$	$\mathbf{W}_2'' = \mathbf{W}_2 \mathbf{D}$
$\mathbf{W}_1' = \mathbf{P}^\mathsf{T} \mathbf{W}_1$	$\mathbf{W}_1'' = \mathbf{D}^{-r} \mathbf{W}_1$
permutation	diagonal

identifiability: are there other (non-equivalent) representations?

Identifiability and geometry

$$\begin{array}{rccc} \varphi = \varphi_{\mathbf{d},\mathbf{r}}: & \mathbb{R}^{\sum d_\ell d_{\ell-1}} & \to & (\mathscr{H}_{d_0,r_1\cdots r_{L-1}})^{d_2} \\ & \theta = (\mathbf{W}_1,\ldots,\mathbf{W}_L) & \mapsto & \mathbf{W}_L \circ \rho_{r_{L-1}} \circ \cdots \circ \rho_{r_1} \circ \mathbf{W}_1 \end{array}$$

homogeneous polynomial function space

Definition. hPNN architecture $(\mathbf{d},\mathbf{r})$ is identifiable

globally :	if φ is	injective	$\varphi^{-1}(\varphi(\theta))/_\sim = \theta$
finitely :		finite-to-one	$\#\varphi^{-1}(\varphi(\theta))/_\sim = c < \infty$

for generic choice of θ

Identifiability and geometry

$$\begin{array}{rccc} \varphi = \varphi_{\mathbf{d},\mathbf{r}}: & \mathbb{R}^{\sum d_\ell d_{\ell-1}} & \to & (\mathscr{H}_{d_0,r_1\cdots r_{L-1}})^{d_2} \\ & \theta = (\mathbf{W}_1,\ldots,\mathbf{W}_L) & \mapsto & \mathbf{W}_L \circ \rho_{r_{L-1}} \circ \cdots \circ \rho_{r_1} \circ \mathbf{W}_1 \end{array}$$

homogeneous polynomial function space

Definition. hPNN architecture $(\mathbf{d},\mathbf{r})$ is identifiable

globally :	if φ is	injective	$\varphi^{-1}(\varphi(\theta))/_\sim = \theta$
finitely :		finite-to-one	$\#\varphi^{-1}(\varphi(\theta))/_\sim = c < \infty$

for generic choice of θ

Proposition (how algebraic geometry helps)

hPNN finitely identifiable

$\Updownarrow$

the neuromanifold (image of φ) has maximal (expected) dimension

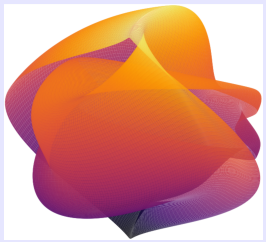
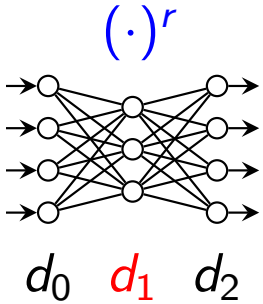
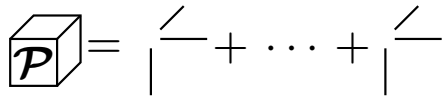


Figure: *V. Shahverdi*

Shallow hPNN and tensors





[Kileel et al. 2019], [Comon et al. 2008] :

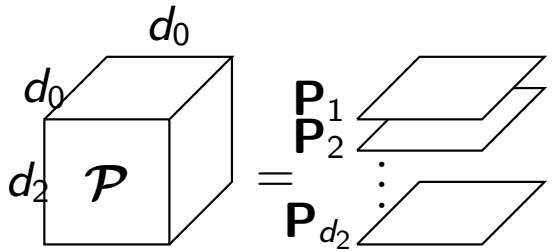
2-layer hPNN
$\mathbf{p}(\mathbf{x}) = \mathbf{W}_2\rho_r(\mathbf{W}_1\mathbf{x})$,
with d_1 hidden neurons

$\leftrightarrow$

partially symmetric
$(r+1)$-order tensor $\mathcal{P}$
of CPD rank d_1

Example: quadratic ($r = 2$)

$$\mathbf{p}(\mathbf{x}) = [\mathbf{x}\mathbf{P}_1\mathbf{x}^\top, \ldots, \mathbf{x}\mathbf{P}_{d_2}\mathbf{x}^\top] \quad \leftrightarrow$$



order-3 tensor

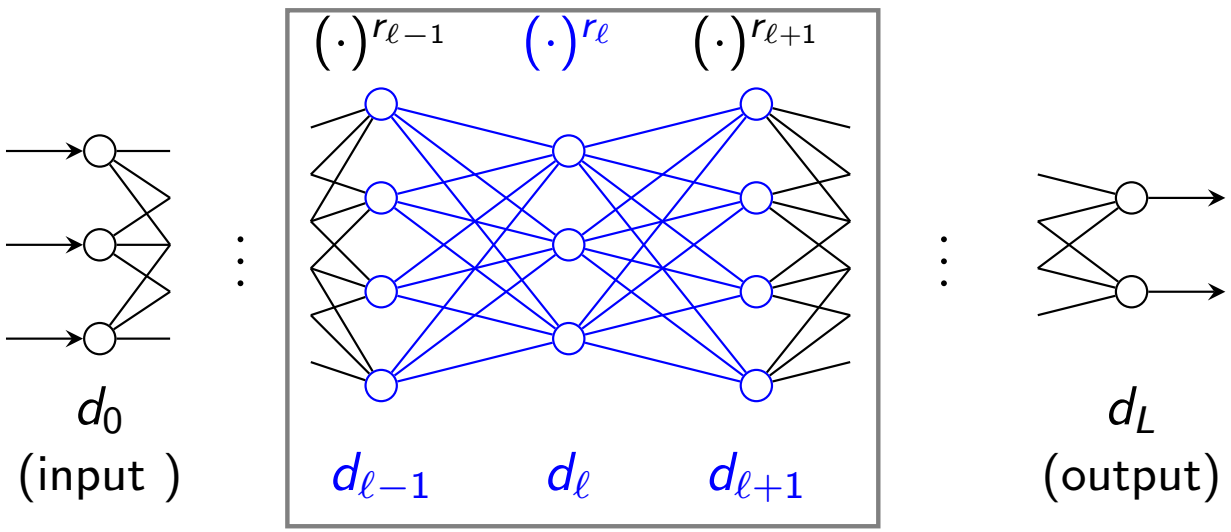
Identifiability of CPD well studied ([Kruskal, 1977], [Domanov, De Lathauwer, 2014], [Chiantini et al., 2014], ...) **Can we leverage it to the deep case?**

Necessary condition for deep case

Identifiability of 2-layer sub-blocks $\Leftarrow$ identifiability of deep PNN

unique representation of $f = \mathbf{W}_L \circ \rho_{r_{L-1}} \circ \mathbf{W}_{L-1} \circ \rho_{r_{L-2}} \circ \cdots \circ \rho_{r_1} \circ \mathbf{W}_1$

$\Rightarrow$ every sub-block $\mathbf{W}_{\ell+1} \circ \rho_{r_\ell} \circ \mathbf{W}_\ell$, has unique representation



L-layer hPNN identifiable $\Rightarrow$ hPNN $(d_{\ell\text{-}1}, d_\ell, d_{\ell+1}), (r_\ell)$ identifiable $\forall \ell$

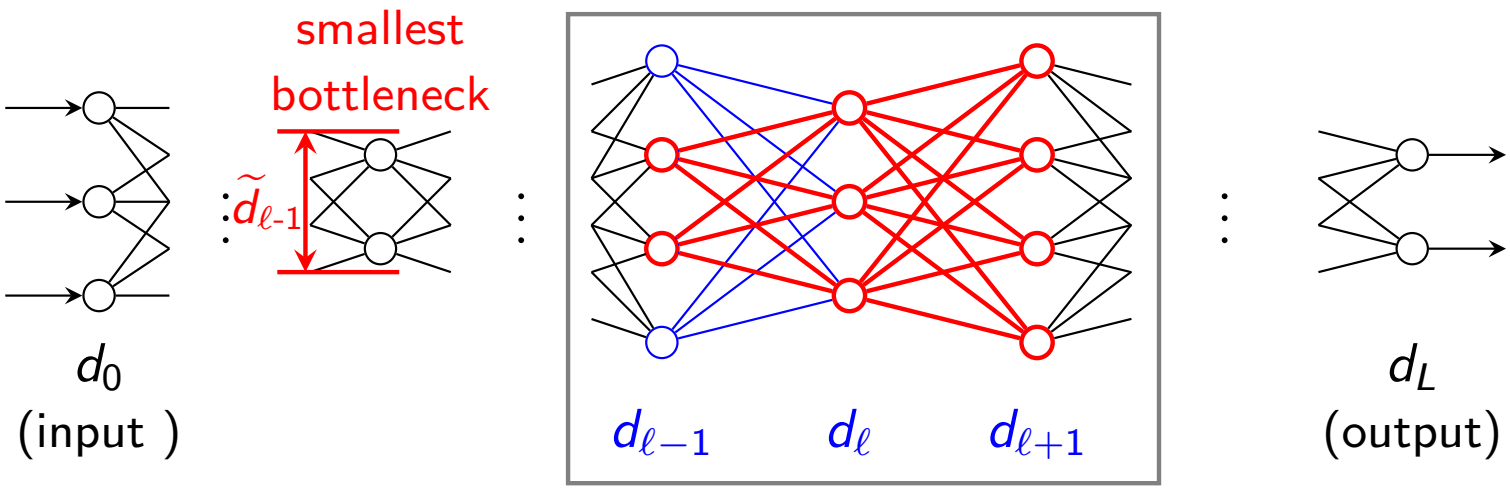
The converse is not true!

Sufficient condition for deep case

Identifiability of 2-layer sub-blocks $\Rightarrow$ identifiability of deep PNN

Theorem (localization of identifiability)

Denote $\tilde{d}_\ell := \min\{d_0, \ldots, d_\ell\}$ *for* $\ell = 0, \ldots, L-2$.
If hPNN architecture $((\tilde{d}_{\ell-1}, d_\ell, d_{\ell+1}), r_\ell)$ *is finitely identifiable* $\forall \ell < L$
then deep hPNN $(d_0, \ldots, d_L), (r_1, \ldots, r_{L-1})$ *is also finitely identifiable.*

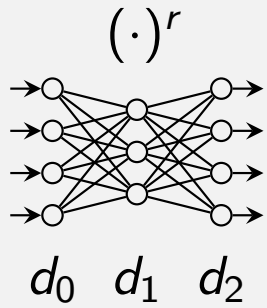


Consequences for different architectures (i)

Weakest 2-layer identifiability condition (Kruskal theorem for CPD)

Proposition (corollary of [Sidiropoulos, Bro, 2000]). If

$$r \geq \frac{2d_1 - \min(d_2, d_1)}{\min(d_1, d_0) - 1},$$



then hPNN architecture $((d_0, d_1, d_2), r)$ is (globally) identifiable.

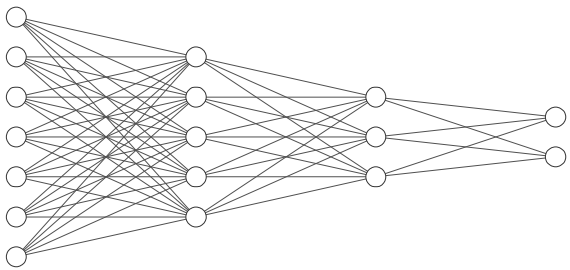
Corollary ((i) linear activation degree thresholds)

Given any widths $(d_0, \ldots, d_L)$, $d_\ell \geq 2$, *hPNN finitely identifiable if*

$$r_\ell \geq 2d_\ell - 1, \quad \forall \ell < L\,.$$

Improves the best algebro-geometric bound $O(d_\ell^2)$ [Finkel et al, 2025]

Consequences for different architectures (ii)



Corollary ((ii) pyramidal architecture)

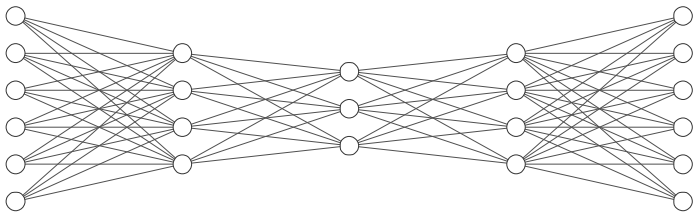
The hPNN with non-increasing widths (except output layer)

$$d_0 \geq d_1 \geq \cdots d_{L-1} \geq 2, \;\; d_L \geq 2$$

is finitely identifiable for $r_\ell \geq 2$.

Settles (and generalizes) conjecture from [Kubjas et al., 2024]

Consequences for different architectures (iii)



Corollary ((iii) bottleneck architecture)

The "bottleneck" hPNN architecture with

$$d_0 \geq d_1 \geq \cdots \geq \ d_b \ \leq d_{b+1} \leq \ldots \leq d_L$$

and $d_b \geq 2$, $r_1, \ldots, r_b \geq 2$, *is finitely identifiable if* $r_\ell \geq \frac{d_\ell}{d_b - 1}$ *in the decoder part* ($\ell > b$) .

decoder should be more complex than encoder

Tackling biases: homogenization

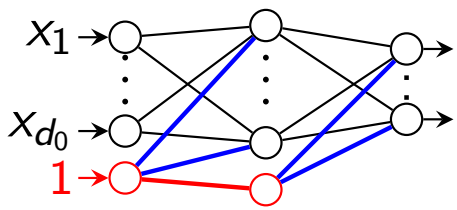
Definition (homogenization — algebraic geometry)

polynomial in d variables $\mapsto$ homogeneous polynomial in $d+1$ variables

Example. $\underbrace{x_1^2 + x_1^2 + x_1 + 3x_2 + 2}_{p(x_1,x_2)} \mapsto \underbrace{x_1^2 + x_1^2 + x_1x_3 + 3x_2x_3 + x_3^2}_{\widetilde{p}(x_1,x_2,x_3)}$

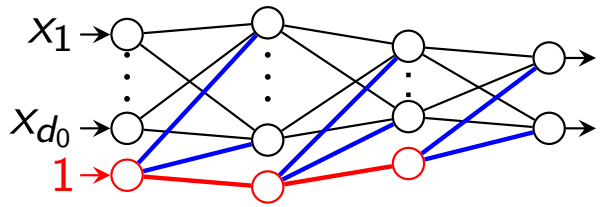
2-layer PNN: $\mathbf{W}_2\rho_r(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2 = \begin{bmatrix}\mathbf{W}_2 & \mathbf{b}_2\end{bmatrix} \rho_r\left(\begin{bmatrix}\mathbf{W}_1 & \mathbf{b}_1 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{x} \\ 1\end{bmatrix}\right)$

PNN (d_0, d_1, d_2)	$\mapsto$	hPNN $(d_0{+}1, d_1{+}1, d_2)$



Identifiability of L-layer PNN with bias

PNN		hPNN
$(d_0, \ldots, d_L)$	$\mapsto$	$(d_0{+}1, \ldots, d_{L-1}{+}1, d_L)$



$$\widetilde{\mathbf{W}}_\ell = \begin{bmatrix} \mathbf{W}_\ell & \mathbf{b}_\ell \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{(d_\ell+1)\times(d_{\ell-1}+1)}, \quad \widetilde{\mathbf{W}}_L = \begin{bmatrix} \mathbf{W}_L & \mathbf{b}_L \end{bmatrix} \in \mathbb{R}^{d_L\times(d_{L-1}+1)}$$

Proposition (Localization of identifiability, bias case)

Let $\widetilde{d}_\ell = \min\{d_0, \ldots, d_\ell\}$.
If 2-layer hPNN $((\tilde{d}_{\ell-1}{+}1, d_\ell{+}1, d_{\ell+1}), r_\ell)$ *is finitely identifiable* $\forall \ell$,
then the L-layer PNN $(\mathbf{d}, \mathbf{r})$ *is finitely identifiable as well.*

- L-layer PNN identifiable if $\quad r_\ell \geq \dfrac{2(d_\ell+1) - \min(d_\ell+1, d_{\ell+1})}{\min(d_\ell, \tilde{d}_{\ell-1})}$
- same corollaries for pyramidal, bottleneck, activation thresholds
- identifiability with single hidden neurons

Summary

- Finite identifiability: shallow $\rightarrow$ deep PNNs
- PNN with biases: homogenization
- Identifiability of several architectures (pyramidal, bottleneck)
- Activation degrees linear in layer widths suffice

- preprint: `https://arxiv.org/abs/2506.17093`
- poster: #4008 (Thu 4 Dec, 11am-2pm)

Thank you!