Neuroalgebraic Geometry

Kathlén Kohn

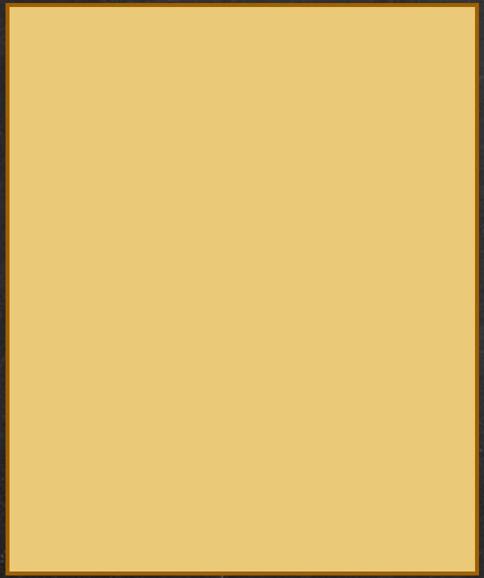
WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM
digital futures
KTH VETENSKAP OCH KONST
SFSG SWEDISH FOUNDATIONS' STARTING GRANT
Ragnar Söderbergs STIFTELSE
Göran Gustafssons Stiftelser

Why does deep learning work?

----

a new and complementary approach

deep learning in a nutshell



parameter space

function space

Goal:
find θ that minimize

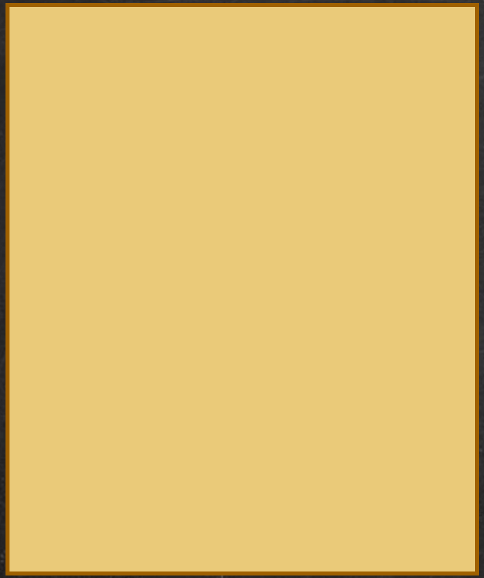
θ ⟼ f_θ ⟼ $L(f_\theta)$

θ
learnable weights

f_θ
network function

$L(f_\theta)$
loss value

deep learning in a nutshell



neuromanifold (Amari et. al 2001)

extremely hard to understand!

parameter space

function space

Goal:

find θ that minimize

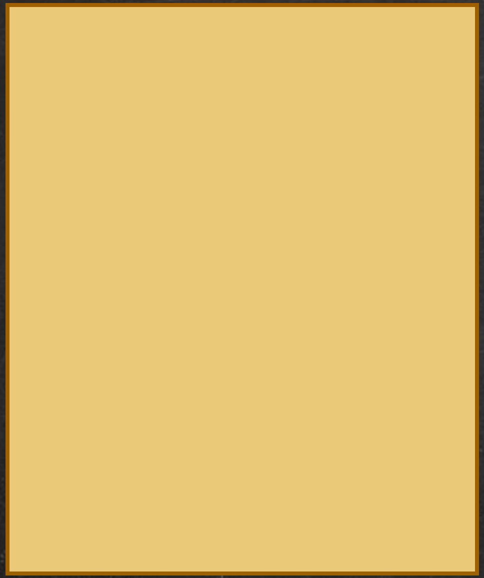
θ $\longmapsto$ f_θ $\longmapsto$ $L(f_\theta)$

learnable weights

network function

loss value

deep learning in a nutshell



parameter space

function space

neuromanifold (Amari et. al 2001)

extremely hard to understand!

Idea:
study polynomial networks

Goal:
find θ that minimize

θ
learnable weights

$\longmapsto$

f_θ
network function

$\longmapsto$

$L(f_\theta)$
loss value

Polynomial neural networks strike an amazing balance

parameter space $\longrightarrow$ neuromanifold

θ
learnable weights $\longmapsto$ f_θ
network function

since they are:

Tractable

- ✓ Polynomial networks are the only ones whose neuromanifold lives in a finite-dimensional vector space [univ. appr. thm.]
- ✓ Can use algebraic geometry

General

- ✓ Arbitrary network can be approximated with polynomial ones

examples

multilayer perceptron (MLP) or **convolutional neural network (CNN)**:

$$\alpha_L \circ \sigma \circ \ldots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$$

α_i: learnable affine linear functions

σ: nonlinear activation function

If σ is a polynomial, then the network is polynomial!

examples

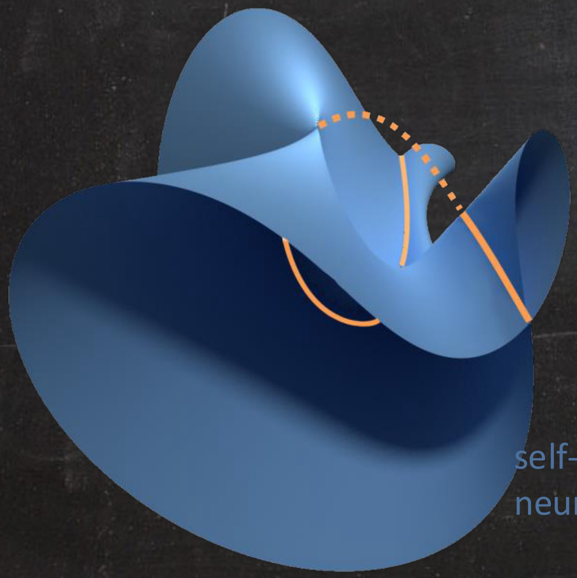
multilayer perceptron (MLP) or **convolutional neural network (CNN)**:

$$\alpha_L \circ \sigma \circ \ldots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$$

α_i: learnable affine linear functions

σ: nonlinear activation function

If σ is a polynomial, then the network is polynomial!



The core of a transformer network is the **self-attention mechanism** that – in its unnormalized version – is a cubic function in the entries of a matrix X:

$$X \longmapsto V X X^T K^T Q X$$

existing approach:

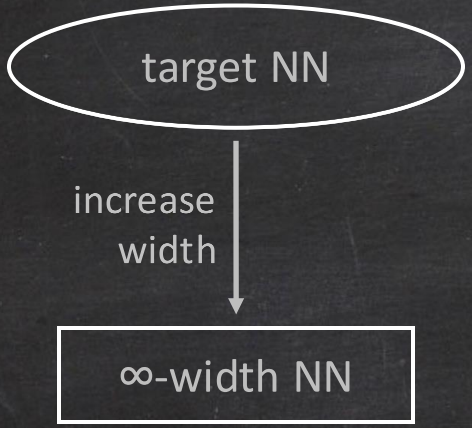
neural tangent kernel

new approach:

algebraic neural network theory

existing approach:

neural tangent kernel



new approach:

algebraic neural network theory

existing approach:

neural tangent kernel

target NN

increase width

∞-width NN

studies linearized models in ∞-dimensional ambient space

aims to draw conclusions **from** the limit

new approach:

algebraic neural network theory

existing approach:

neural tangent kernel

$$\text{target NN} \xrightarrow{\text{increase width}} \infty\text{-width NN}$$

studies linearized models
in ∞-dimensional ambient space

aims to draw conclusions
from the limit

new approach:

algebraic neural network theory

$$\text{algebraic NNs} \xrightarrow{\text{polynomial approximation}} \text{target NN}$$

existing approach:

neural tangent kernel

target NN

increase width

∞-width NN

studies linearized models
in ∞-dimensional ambient space

aims to draw conclusions
from the limit

new approach:

algebraic neural network theory

algebraic NNs

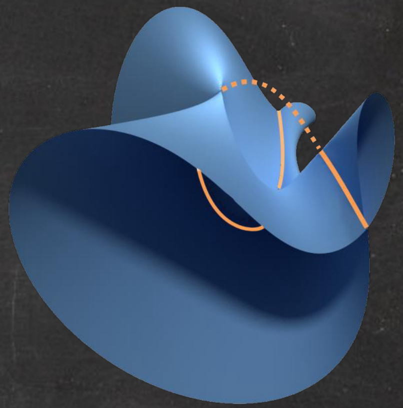
polynomial approximation

target NN

studies nonlinear models
in finite-dimensional ambient space

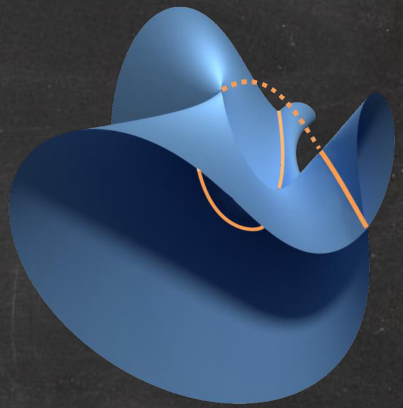
aims to draw conclusions
in the limit

singularities of neuromanifolds

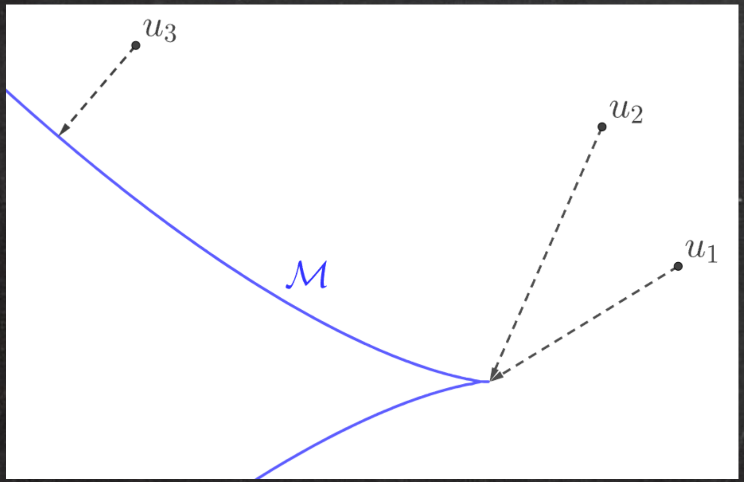


- ☹ can cause numerical instability and slow leaning (Amari et. al)
- ☺ and solutions that generalize better (Kohn et. al)

singularities of neuromanifolds

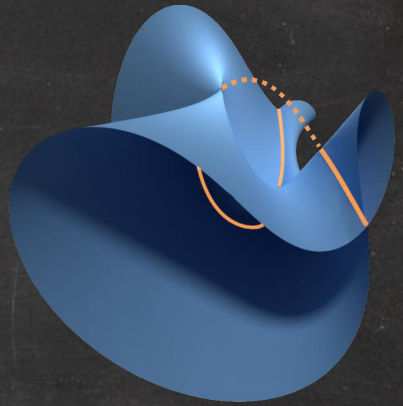


- ☹ can cause numerical instability and slow leaning (Amari et. al)
- ☺ and solutions that generalize better (Kohn et. al)

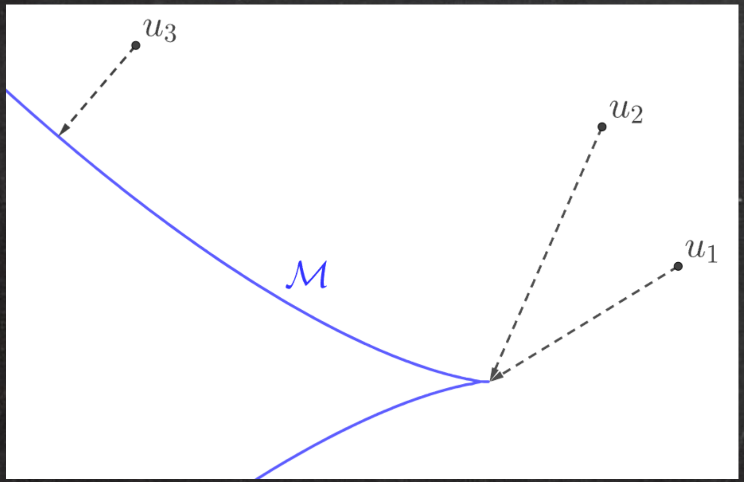


the singularity is a stable solution

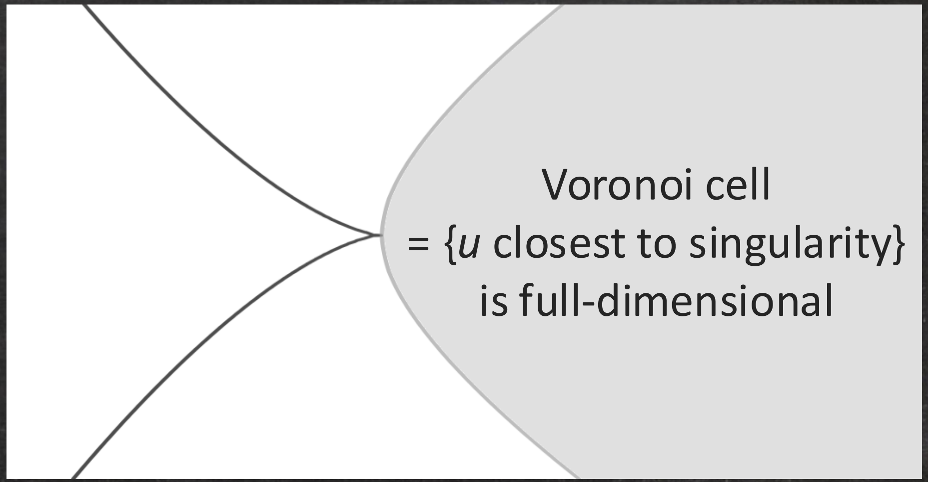
singularities of neuromanifolds



- ☹ can cause numerical instability and slow leaning (Amari et. al)
- ☺ and solutions that generalize better (Kohn et. al)

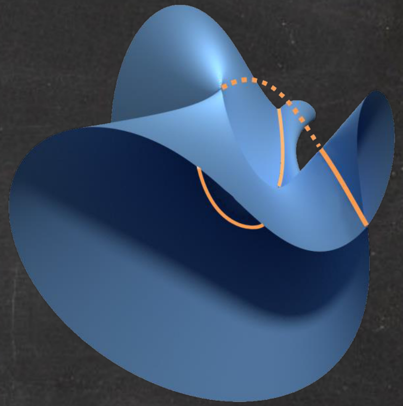


the singularity is a stable solution

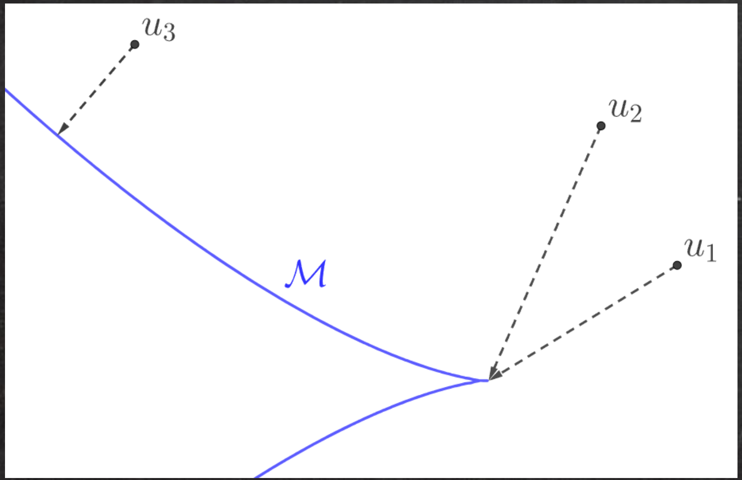


=> implicit bias to singularity

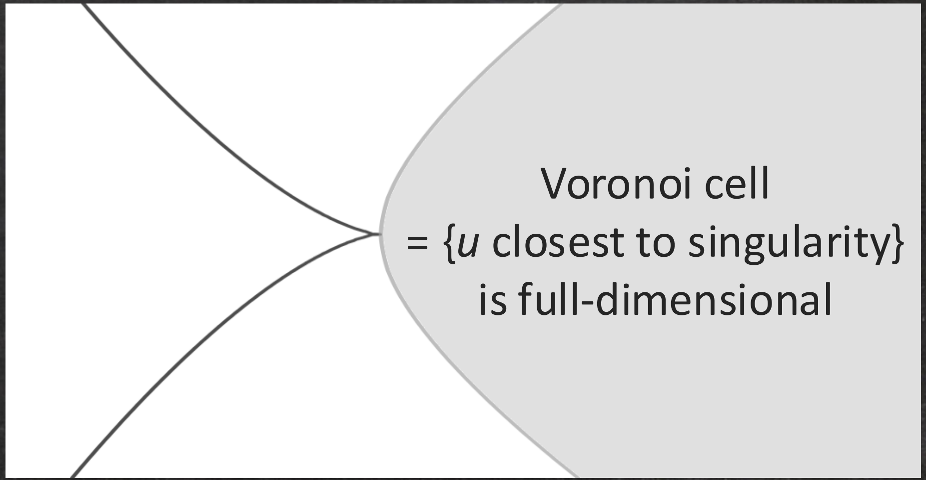
singularities of neuromanifolds



- can cause numerical instability and slow leaning (Amari et. al)
- and solutions that generalize better (Kohn et. al)



the singularity is a stable solution



=> implicit bias to singularity

Conjecture: Singularities of neuromanifold are sparse subnetworks (proven for MLPs and CNNs).

Example: MLPs

multilayer perceptrons

$$\alpha_L \circ \sigma \circ \dots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$$

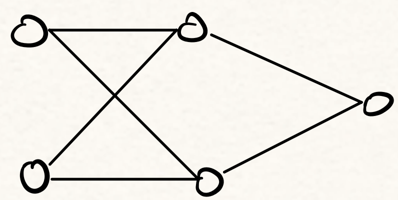
α_i = <u>learnable</u> affine linear functions

σ = nonlinear activation function, applied entrywise

we assume: σ is a univariate polynomial

Ex: $\sigma(x) = x^2$

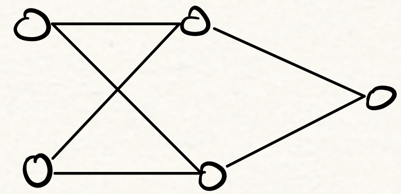
$$\begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right)$$



Which functions does this MLP parametrize?

Ex.: $\sigma(x) = x^2$

$$\begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$



Which functions does this MLP parametrize?

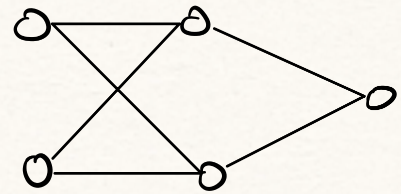
$$e(ax+by)^2 + f(cx+dy)^2$$
$$= \underbrace{(a^2e + c^2f)}_{A} x^2 + \underbrace{2(abe + cdf)}_{B} xy + \underbrace{(b^2e + d^2f)}_{C} y^2$$

Can you obtain <u>all</u> of $\mathbb{R}[x,y]_2$?

↖ homogeneous quadratic polynomials in x,y

i.e., are all values for A, B, C possible?

Ex.: $\sigma(x) = x^2$



$$\begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

Which functions does this MLP parametrize?

$$e(ax+by)^2 + f(cx+dy)^2$$
$$= \underbrace{(a^2e + c^2f)}_{A} x^2 + \underbrace{2(abe + cdf)}_{B} xy + \underbrace{(b^2e + d^2f)}_{C} y^2$$

Can you obtain <u>all</u> of $\mathbb{R}[x,y]_2$?

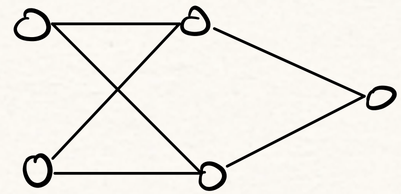
↖ homogeneous quadratic polynomials in x,y

i.e., are all values for A, B, C possible?

YES

What about $\sigma(x) = x^3$?

Ex.: $\sigma(x) = x^3$



$$\begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right)$$

Which functions does this MLP parametrize?

$$e(ax+by)^3 + f(cx+dy)^3$$
$$= \underbrace{(a^3e+c^3f)}_{A}x^3 + \underbrace{3(a^2be+c^2df)}_{B}x^2y + \underbrace{3(ab^2e+cd^2f)}_{C}xy^2 + \underbrace{(b^3e+d^3f)}_{D}y^3$$

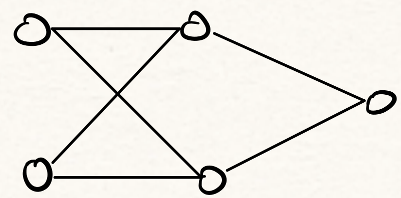
Can you obtain <u>all</u> of $\mathbb{R}[x,y]_3$?

↖ homogeneous cubic polynomials in x,y

i.e., are all values for A,B,C,D possible?

Ex.: $\sigma(x) = x^3$

$$[e \;\; f]\, \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$



Which functions does this MLP parametrize?

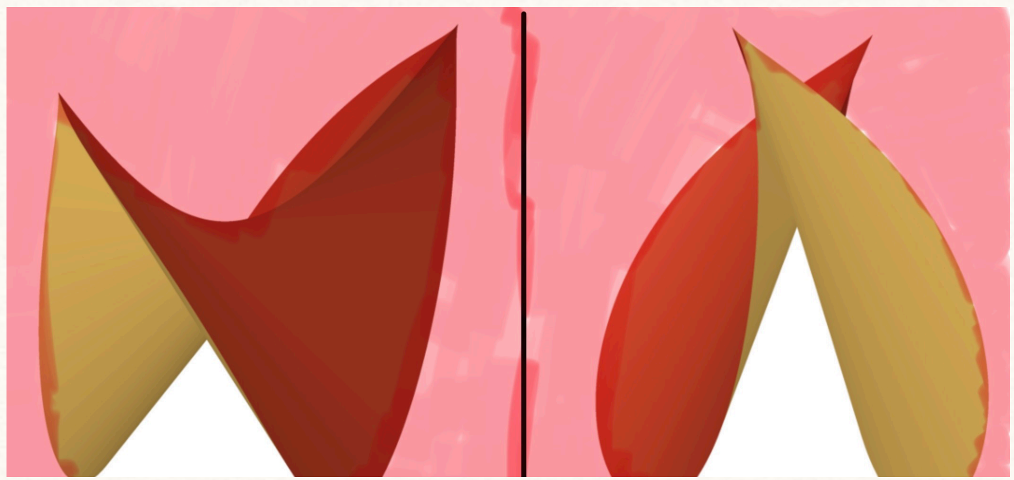
$$e(ax+by)^3 + f(cx+dy)^3$$
$$= \underbrace{(a^3e + c^3f)}_{A}x^3 + \underbrace{3(a^2be + c^2df)}_{B}x^2y + \underbrace{3(ab^2e + cd^2f)}_{C}xy^2 + \underbrace{(b^3e + d^3f)}_{D}y^3$$

Can you obtain <u>all</u> of $\mathbb{R}[x,y]_3$?

↖ homogeneous cubic polynomials in x,y

i.e., are all values for A, B, C, D possible?

No, e.g. $A = 1$, $B = 0$, $C = -1$, $D = 0$



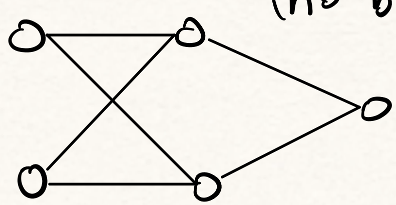
Neuromanifolds

A **parametric machine learning** model is a map $\mu: \Theta \times X \to Y$.

parameters inputs outputs

Its **neuromanifold** is $\mathcal{M} := \{\mu(\theta,\cdot): X \to Y \mid \theta \in \Theta\}$.

Example MLPs:



(no bias vectors = linear layer functions)

$\sigma(x) = x^2 \quad \Rightarrow \quad \mathcal{M} = \mathbb{R}[x,y]_2$

$\sigma(x) = x^3 \quad \Rightarrow \quad \mathcal{M} \subsetneq \mathbb{R}[x,y]_3$

Polynomial MLPs: $\alpha_L \circ \sigma \circ \dots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$, where

$\alpha_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ affine linear

$\sigma \in \mathbb{R}[x]_{\leq \delta}$

$\Rightarrow$ $\mathcal{M}$ lives in a <u>finite-dimensional</u> vector space, namely

$$\left(\mathbb{R}[x_1, \dots, x_{d_0}]_{\leq \delta^{L-1}}\right)^{d_L}$$

Polynomial MLPs: $\alpha_L \circ \sigma \circ \ldots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$, where

$\alpha_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ affine linear

$\sigma \in \mathbb{R}[x]_{\leq \delta}$

$\Rightarrow \mathcal{M}$ lives in a finite-dimensional vector space, namely

$$\left(\mathbb{R}[x_1, \ldots, x_{d_0}]_{\leq \delta^{L-1}}\right)^{d_L}$$

Polynomial MLPs are the only ones with that property!

Universal Approximation Theorem

Leshno, Lin, Pinkus, Schocken: Multilayer feedforward networks with a non-polynomial activation function can approximate any function.
Neural Networks 6, 1993:

Theorem 1:

Let $\sigma \in M$. Set

$$\Sigma_n = \text{span}\,\{\sigma(\mathbf{w} \cdot \mathbf{x} + \theta) : \mathbf{w} \in R^n, \theta \in R\}.$$

Then Σ_n is dense in $C(R^n)$ if and only if σ is not an algebraic polynomial (a.e.).

Universal Approximation Theorem

Leshno, Lin, Pinkus, Schocken: Multilayer feedforward networks with a non-polynomial activation function can approximate any function.
Neural Networks 6, 1993:

Theorem 1:

Let $\sigma \in M$. Set

$$\Sigma_n = \operatorname{span}\{\sigma(\mathbf{w}\cdot\mathbf{x}+\theta) : \mathbf{w} \in R^n, \theta \in R\}.$$

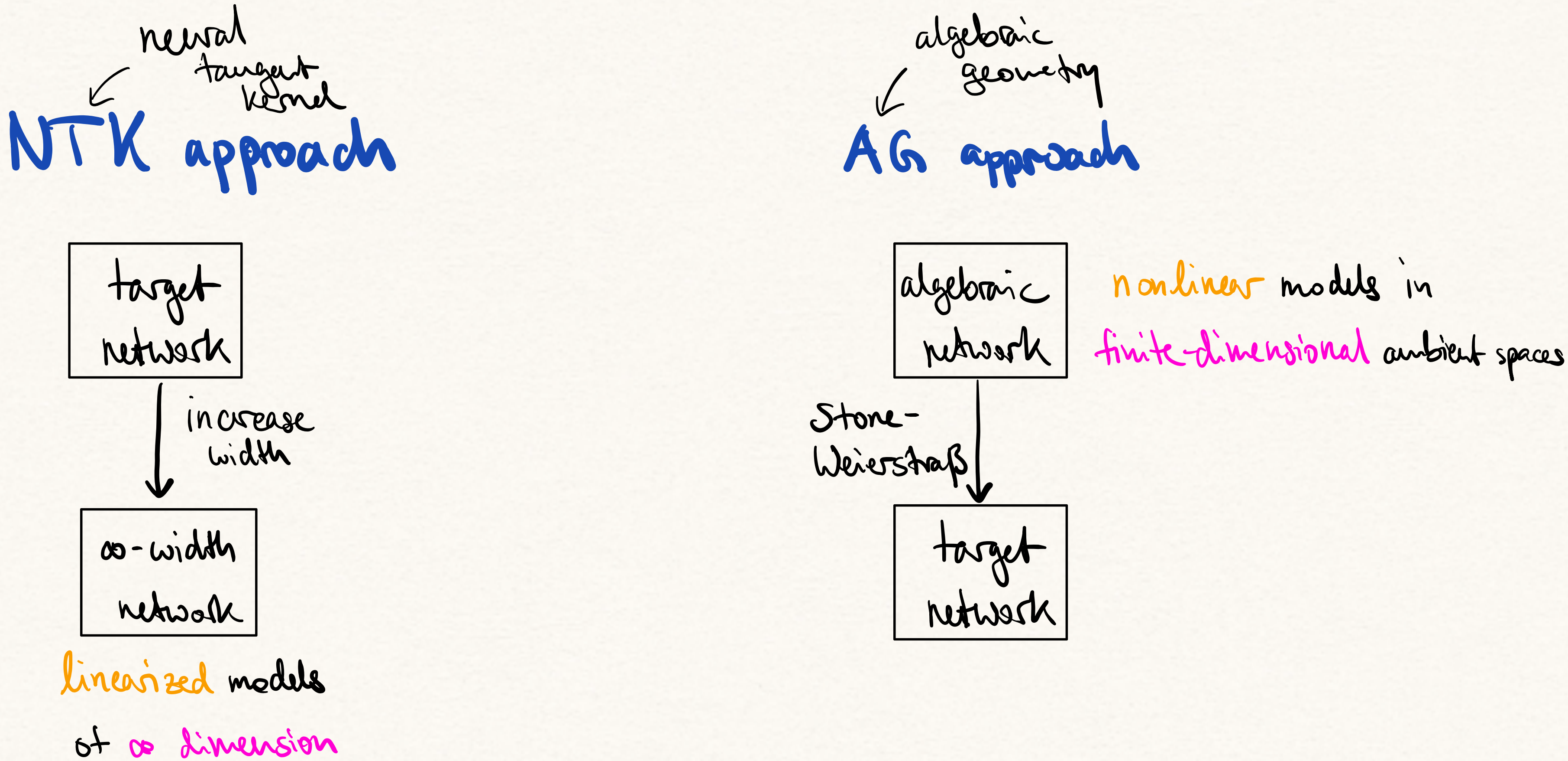
Then Σ_n is dense in $C(R^n)$ if and only if σ is not an algebraic polynomial (a.e.).

polynomials are <u>the</u> choice to approximate networks with finite-dimensional models

AG approach

$$\begin{tikzcd}
\boxed{\text{algebraic network}} \arrow[d, "\text{Stone-Weierstraß}"'] \\
\boxed{\text{target network}}
\end{tikzcd}$$

nonlinear models in finite-dimensional ambient spaces



Neural Tangent Kernel: Convergence and Generalization in Neural Networks

Arthur Jacot, Franck Gabriel, Clement Hongler

Advances in Neural Information Processing Systems 31 (NeurIPS 2018)

Network training = 'distance' minimization

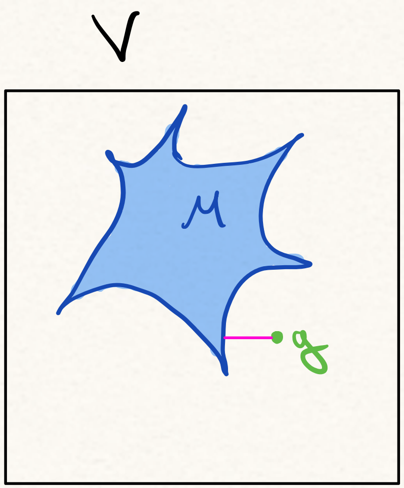
Let $\mathcal{M} \subseteq V := \left(\mathbb{R}[x_1,\dots,x_{d_0}]_{\leq D}\right)^{d_L}$,

neuromanifold

$S \subseteq \mathbb{R}^{d_0} \times \mathbb{R}^{d_L}$ finite dataset,

mean squared error

MSE loss: $\mathcal{L}(f) := \sum_{(a,b)\in S} \|f(a)-b\|^2$



Proposition:

[$\mathrm{dist}(f,g) = 0$ possible for $f \neq g$]

There is a pseudometric $\mathrm{dist}: V \times V \to \mathbb{R}_{\geq 0}$ and some $g \in V$ such that minimizing $\mathcal{L}(f)$ over $f \in \mathcal{M}$ is equivalent to minimizing $\mathrm{dist}(f,g)$ over $f \in \mathcal{M}$.

Loss Landscape

$$= \{(\theta, \mathcal{L}(f_\theta)) \mid \theta \in \Theta\}$$

params, input, output

$$\mu : \underset{\text{params}}{\Theta} \times \underset{\text{input}}{X} \to \underset{\text{output}}{Y}$$

$$\mathcal{M} = \{\underbrace{\mu(\theta, \cdot)}_{=f_\theta} \mid \theta \in \Theta\}$$

Loss Landscape

$$= \{(\theta, \mathcal{L}(f_\theta)) \mid \theta \in \Theta\}$$

$$\mu : \underset{\text{params}}{\Theta} \times \underset{\text{input}}{X} \to \underset{\text{output}}{Y}$$

$$\mathcal{M} = \{\underbrace{\mu(\theta, \cdot)}_{= f_\theta} \mid \theta \in \Theta\}$$

can be studied in a decoupled way:

$$\begin{array}{ccccc} \Theta & \longrightarrow & \mathcal{M} & \xrightarrow{\mathcal{L}} & \mathbb{R} \\ \theta & \longmapsto & f_\theta & & \end{array}$$

$$\rightsquigarrow$$

loss landscape in function space:

$$= \{(f, \mathcal{L}(f)) \mid f \in \mathcal{M}\} \subseteq V \times \mathbb{R}$$

Loss Landscape

$$= \{(\theta, \mathcal{L}(f_\theta)) \mid \theta \in \Theta\}$$

$$\mu : \overset{\text{params}}{\Theta} \times \overset{\text{input}}{X} \to \overset{\text{output}}{Y}$$

$$\mathcal{M} = \{ \underbrace{\mu(\theta, \cdot)}_{= f_\theta} \mid \theta \in \Theta \}$$

can be studied in a decoupled way:

$$\begin{tikzcd}
\Theta \arrow[r] & \mathcal{M} \arrow[r, "\mathcal{L}"] & \mathbb{R} \\
\theta \arrow[r, mapsto] & f_\theta &
\end{tikzcd}$$

loss landscape in function space:

$$= \{(f, \mathcal{L}(f)) \mid f \in \mathcal{M}\} \subseteq V \times \mathbb{R}$$

Geometry of $\mathcal{M}$ affects loss landscape!

How?

Which geometric properties does $\mathcal{M}$ have?

Geometry of Neuromanifolds

$\mu : \Theta \times X \longrightarrow Y$ polynomial (in both $\theta \in \Theta$ & $x \in X$)

$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta, \cdot) \end{aligned}$$

What kind of object is $\mathcal{M}$?

Geometry of Neuromanifolds

$\mu : \Theta \times X \longrightarrow Y$ polynomial (in both $\theta \in \Theta$ & $x \in X$)

$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta, \cdot) \end{aligned}$$

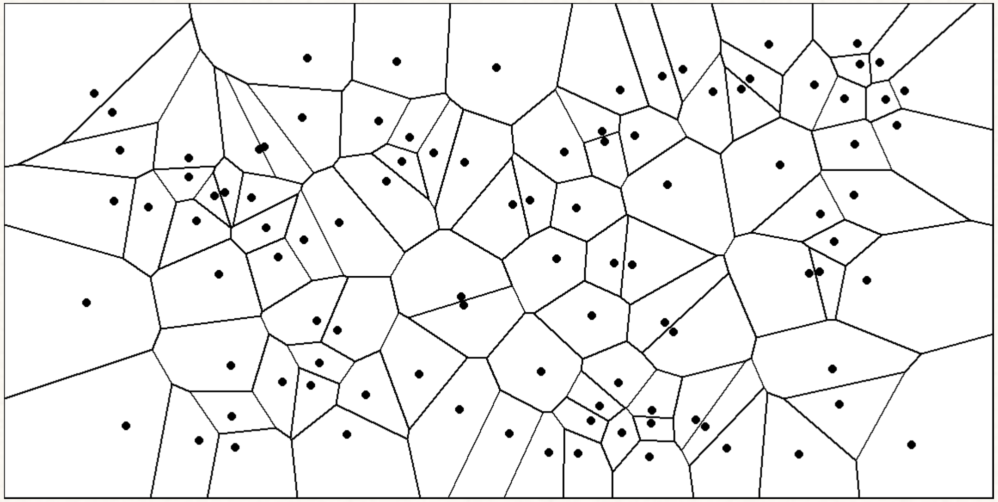
What kind of object is $\mathcal{M}$?

A **semialgebraic** set!

describable by polynomial equations & inequalities

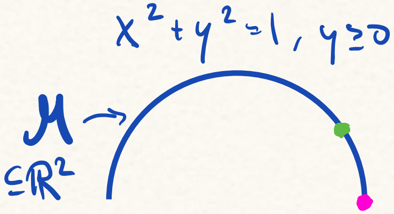
Euclidean distance minimization can be implicitly biased to singularities & boundaries of $\mathcal{M}$

Voronoi cells



For $S \subseteq \mathbb{R}^n$, the **Voronoi cell** at $p \in S$ is

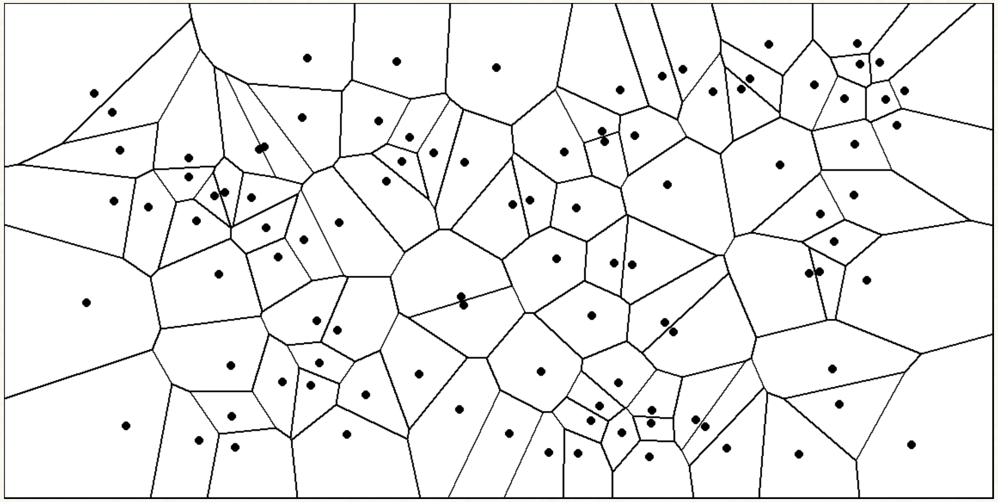
$$\mathrm{Vor}_S(p) := \{ u \in \mathbb{R}^n \mid \forall q \in S,\ q \neq p : \|p-u\|_2 < \|q-u\|_2 \}$$



What is the Voronoi cell at $\bullet$?

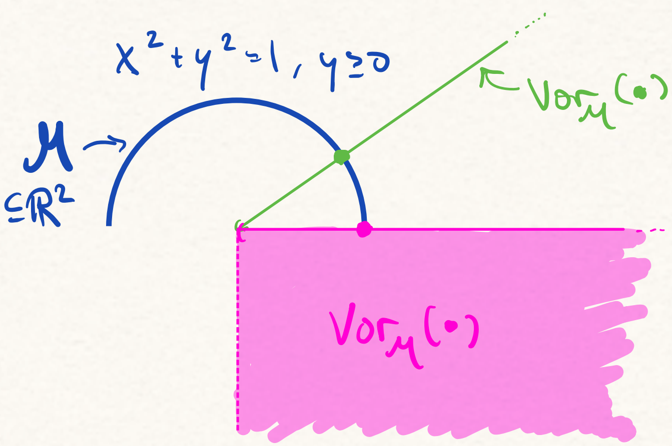
What is the Voronoi cell at $\bullet$?

Voronoi cells



For $S \subseteq \mathbb{R}^n$, the **Voronoi cell** at $p \in S$ is

$$\mathrm{Vor}_S(p) := \{ u \in \mathbb{R}^n \mid \forall q \in S,\ q \neq p: \|p-u\|_2 < \|q-u\|_2 \}$$

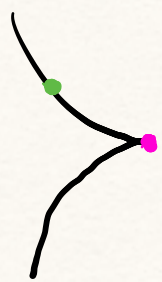


The 2 relative boundary points are the only points on $\mathcal{M}$ with full-dimensional Voronoi cells!

$\rightsquigarrow$ **implicit bias** towards $\partial\mathcal{M}$

points in $\partial\mathcal{M}$ are global minima with positive probability on data u

singularities

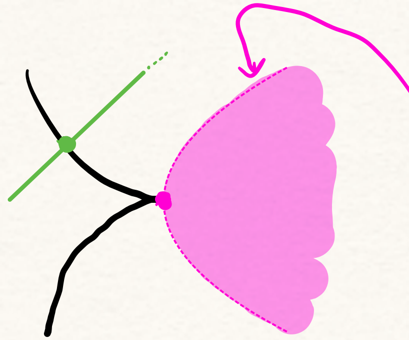


What are the Voronoi cells at • and • ?

singularities

$y^2 + x^3 = 0$

$t \mapsto (-t^2, t^3)$



Challenge: Compute this curve!

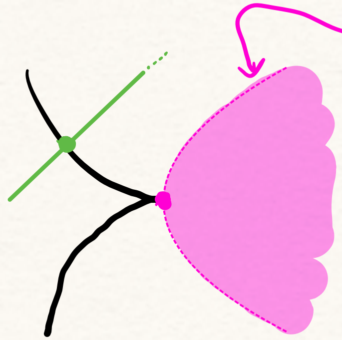
$\rightsquigarrow$ **implicit bias** towards $\mathrm{Sing}(M)$

What are the Voronoi cells at • and • ?

singularities

$y^2+x^3=0$

$t \mapsto (-t^2, t^3)$



Challenge: Compute this curve!

$\rightsquigarrow$ **implicit bias** towards $\mathrm{Sing}(\mathcal{M})$

What are the Voronoi cells at • and • ?

<u>Tradeoff</u>

☹

learning close to singularity
$\rightsquigarrow$ slow & numerical instability
[Amari et al]

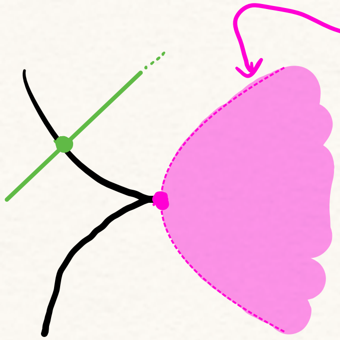
☺

singular solution generalizes better:
① stable global minimum when perturbing data
② **Conjecture:** singularities of neuromanifolds are sparse subnetworks
[we've proven this for MLPs & CNNs]

singularities

$y^2 + x^3 = 0$

$t \mapsto (-t^2, t^3)$



Challenge: Compute this curve!

$\rightsquigarrow$ **implicit bias** towards Sing(M)

What are the Voronoi cells at • and • ?

Tradeoff

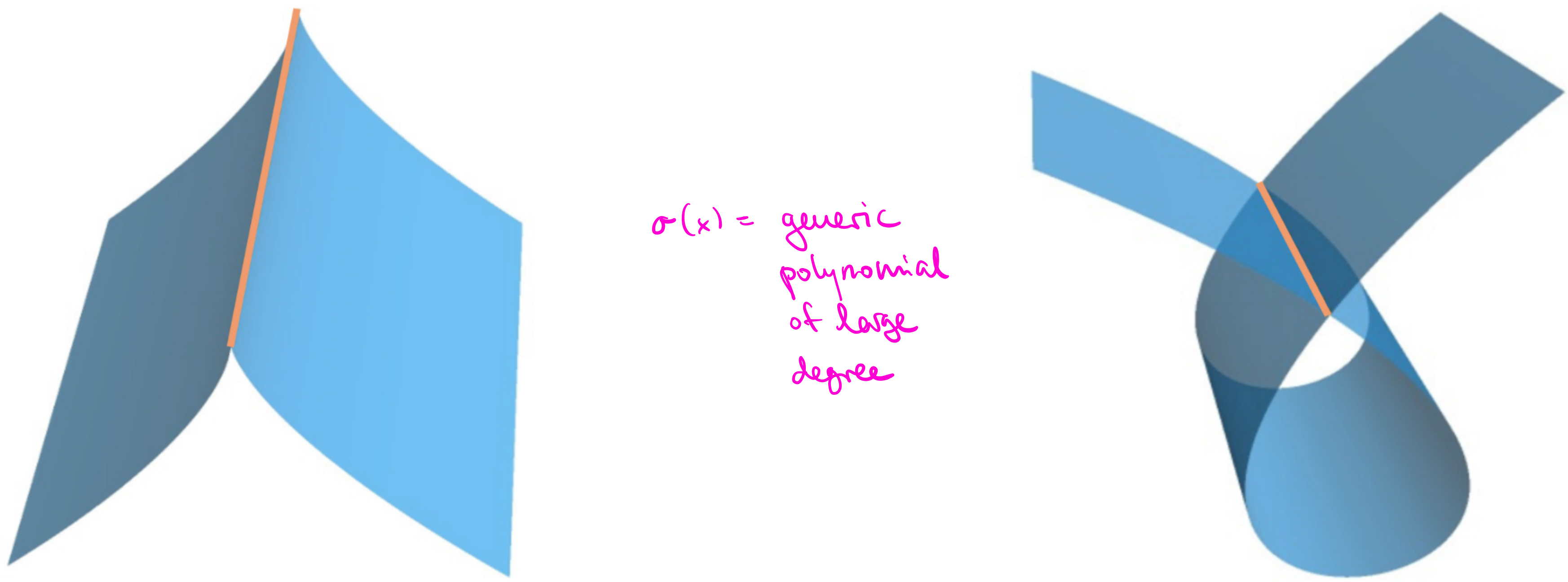
☹

learning close to singularity
$\rightsquigarrow$ slow & numerical instability
[Amari et al]

☺

singular solution generalizes better:
① stable global minimum when perturbing data
② **Conjecture:** singularities of neuromanifolds are sparse subnetworks
[We've proven this for MLPs & CNNs]

In general: depends on **type** of singularity



MLP CNN

$\sigma(x) =$ generic polynomial of large degree

These singularities have that tradeoff, ... while these don't!

In both cases, they are sparse subnetworks :)

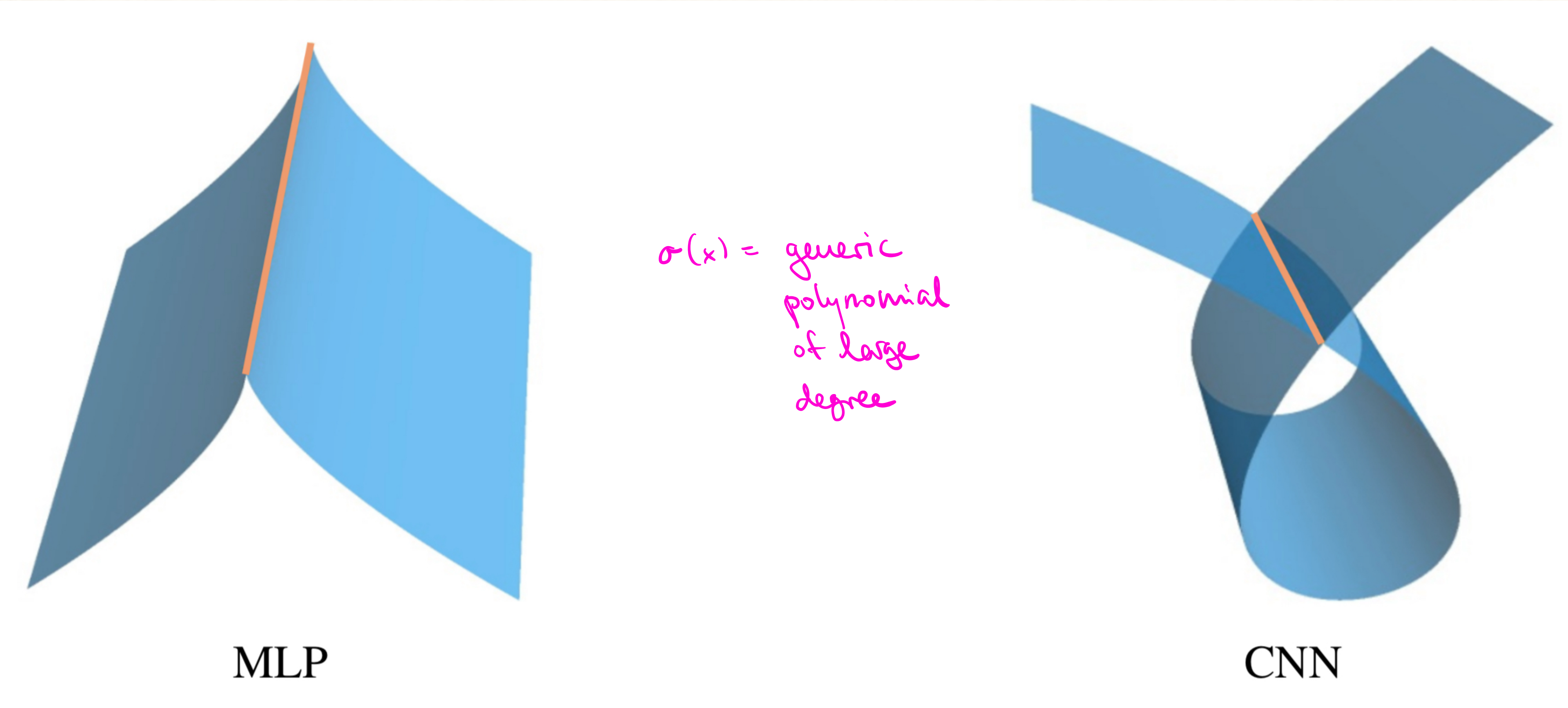
Singularities are (essentially) determined by the

<u>parameter symmetries</u> of the network

$\mu^{-1}(f) := \{ \theta \in \Theta \mid \mu(\theta, \cdot) = f \}$ for $f \in \mathcal{M}$

Singularities are (essentially) determined by the **parameter symmetries** of the network

$\mu^{-1}(f) := \{\theta \in \Theta \mid \mu(\theta, \cdot) = f\}$ for $f \in \mathcal{M}$



$\sigma(x)$ = generic polynomial of large degree

$\mu^{-1}(f)$ = neuron permutations	**for almost all f**	$\lvert\mu^{-1}(f)\rvert = 1$
$\lvert\mu^{-1}(f)\rvert = \infty$	**for subnetworks f**	$1 < \lvert\mu^{-1}(f)\rvert < \infty$

Wallenberg Advanced Scientific Forum

"Symmetries in Neural Networks"

Sept 29 – Oct 2, 2026

Sweden

Interactive problem-solving workshop:

Develop systematic tools to

1. identify & compute symmetries
2. determine their impact on training, generalization, scalability, reliability, ...

2 open slots!

apply with CV & motivation: kathlen@kth.se
by April 30

Erik Bekkers (Univ. of Amsterdam)
Misha Belkin (tentative; UCSD)
Georg Bökman (Univ. of Amsterdam)
Marie Charlotte Brandenburg (Ruhr-U)
Nadav Cohen (Tel Aviv University)
Bella Finkel (Univ. Wisconsin-Madison)
Remi Gribonval (École Normale Supérieure de Lyon)
Moritz Grillo (MPI MiS Leipzig)
Elisenda Grigsby (Boston College)
Stefanie Jegelka (TU Munich and MIT)
Joe Kileel (UT Austin)
Emily King (Colorado State)
Kaie Kubjas (Aalto)
Henry Kvinge (Pacific Northwest National Lab/University of Washington)
Kathryn Lindsey (Boston College)
Guido Montúfar (MPI MiS Leipzig and UCLA)
Jose I. Rodriguez (Univ. Wisconsin-Madison)
Mariia Seleznova (LMU)
Rekha Thomas (Univ. of Washington)
Konstantin Usevich (Nancy)
Rene Vidal (U Penn)
Soledad Villar (John Hopkins)
Robin Walters (Northeastern)
Rachel Ward (UT Austin)
Melanie Weber (Harvard)

Non-Sweden Participants

GeUmetric Deep Learning Workshop 2026

Umeå

August 17-19

register here

Identifiable Equivariant Networks are Layerwise Equivariant

Vahid Shahverdi [*1] **Giovanni Luca Marchetti** [*1] **Georg Bökman** [*2] **Kathlén Kohn** [*1]

Abstract

We investigate the relation between end-to-end equivariance and layerwise equivariance in deep neural networks. We prove the following: For a network whose end-to-end function is equivariant with respect to group actions on the input and output spaces, there is a parameter choice yielding the same end-to-end function such that its layers are equivariant with respect to some group actions on the latent spaces. Our result assumes that the parameters of the model are identifiable in an appropriate sense. This identifiability prop-

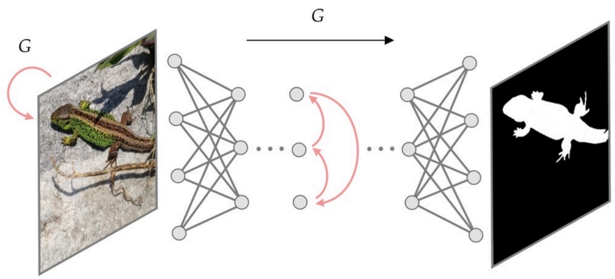
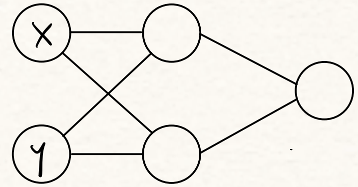


Figure 1. An image segmentation model is equivariant to rotations of the image. Our main result implies that the group action on the input propagates through the network via latent symmetries (e.g., neuron permutations), until it reaches the output.



$$f = \begin{bmatrix} -2 & 0 \end{bmatrix} \sigma\left(\underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}} \right)$$

invariant under S_2, but is <u>not</u> equivariant.

$$f = \begin{bmatrix} -2 & 0 \end{bmatrix} \sigma\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

is layerwise equivariant ☺

Identifiable Equivariant Networks are Layerwise Equivariant

Vahid Shahverdi [*1] **Giovanni Luca Marchetti** [*1] **Georg Bökman** [*2] **Kathlén Kohn** [*1]

Abstract

We investigate the relation between end-to-end equivariance and layerwise equivariance in deep neural networks. We prove the following: For a network whose end-to-end function is equivariant with respect to group actions on the input and output spaces, there is a parameter choice yielding the same end-to-end function such that its layers are equivariant with respect to some group actions on the latent spaces. Our result assumes that the parameters of the model are identifiable in an appropriate sense. This identifiability prop-

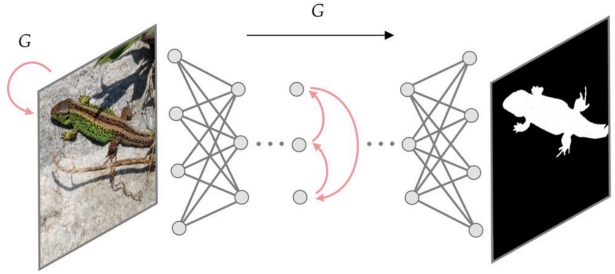
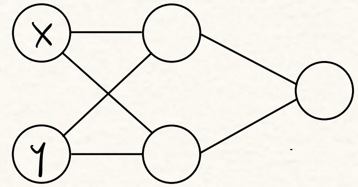


Figure 1. An image segmentation model is equivariant to rotations of the image. Our main result implies that the group action on the input propagates through the network via latent symmetries (e.g., neuron permutations), until it reaches the output.



$$f = \begin{bmatrix} -2 & 0 \end{bmatrix} \sigma\left(\underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}} \right)$$

invariant under S_2, but is <u>not</u> equivariant.

$$f = \begin{bmatrix} -2 & 0 \end{bmatrix} \sigma\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

is layerwise equivariant :)

(a) Autoencoder with Tanh. *(b)* Classifier with Tanh. *(c)* Autoencoder with GELU. *(d)* Classifier with GELU.

Figure 2. First-layer weights of MLPs trained on CIFAR10. Each square is a filter that maps an input RGB image to a single neuron of the subsequent layer. The filters have been sorted for illustrative purposes, with different filter categories highlighted by different colors.

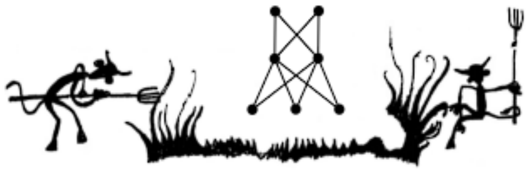
Pink: Filters with a left-to-right mirrored copy (shown adjacent).
Light Blue: Filters with a left-to-right mirrored negated copy (shown adjacent).
Gold: Filters with a negated copy (shown adjacent).
Purple: Left-to-right symmetric.
Dark Blue: Left-to-right anti-symmetric.
Gray: Filters that do not fit into the other categories.

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]

Abstract

In this position paper, we promote the study of function spaces parameterized by machine learning models through the lens of algebraic geometry. To this end, we focus on algebraic models, such as neural networks with polynomial activa-



ICML 2025 Spotlight

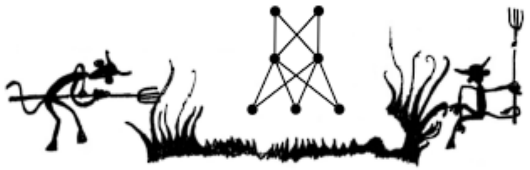
machine learning	algebraic geometry
subnetworks & implicit bias	singularities
sample complexity & expressivity	dimension, degree, covering number
identifiability & invariance	fibers
optimization & gradient descent	critical point theory, discriminants, dynamical invariants

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]

Abstract

In this position paper, we promote the study of function spaces parameterized by machine learning models through the lens of algebraic geometry. To this end, we focus on algebraic models, such as neural networks with polynomial activa-



ICML 2025 Spotlight

machine learning	algebraic geometry
subnetworks & implicit bias	singularities
sample complexity & expressivity	dimension, degree, covering number
identifiability & invariance	fibers
optimization & gradient descent	critical point theory, discriminants, dynamical invariants

Goals:

1. Develop a complete catalog between:

architecture choices	⟺	geometry of neuromanifold	⟺	network training & learning

2. Devise mathematically grounded design strategies for neural networks.

algebraic neural network theory – an emerging field

Kileel, Trager, Bruna: On the expressive power of deep polynomial neural networks. **NeurIPS** 2019.

Trager, Kohn, Bruna: Pure and spurious critical points: a geometric study of linear networks. **ICLR** 2020.

Kohn, Merkh, Montúfar, Trager: Geometry of linear convolutional networks. **SIAM Journal on Applied Algebra & Geometry** 2022.

Kohn, Montúfar, Shahverdi, Trager: Function space & critical points of linear CNNs. **SIAM Journal on Applied Algebra & Geometry** 2024.

Kubjas, Li, Wiesmann: Geometry of polynomial neural networks. **Algebraic Statistics** 2024.

Marchetti, Shahverdi, Mereta, Trager, Kohn: Position: Algebra unveils deep learning – An invitation to Neuroalgebraic Geometry. **ICML** 2025.

Mody, Zubkov: Geometry of Rank Constraints in Shallow Polynomial Neural Networks. **ICML** 2025 Workshop.

Zhang, Kileel: Covering number of real algebraic varieties and beyond: Improved bounds and applications. **FoCM** 2025.

Finkel, Rodriguez, Wu, Yahl: Activation thresholds and expressiveness of polynomial neural networks. **Algebraic Statistics** 2025.

Arjevani, Bruna, Kileel, Polak, Trager: Geometry and optimization of shallow polynomial networks. SIAM J. Applied Algebra & Geometry 2026.

Shahverdi, Marchetti, Kohn: On the geometry and optimization of polynomial convolutional networks. **AISTATS** 2025.

Henry, Marchetti, Kohn: Geometry of lightning self-attention: Identifiability and dimension. **ICLR** 2025.

Massarenti, Mella: The Alexander-Hirschowitz theorem for neurovarieties. arXiv:2511.19703.

Grosdos, Robeva, Zubkov: Algebraic geometry of rational neural networks. arXiv:2509.11088.

Shahverdi, Marchetti, Kohn: Learning on a razor’s edge: the singularity bias of polynomial neural networks. **ICLR 2026**.

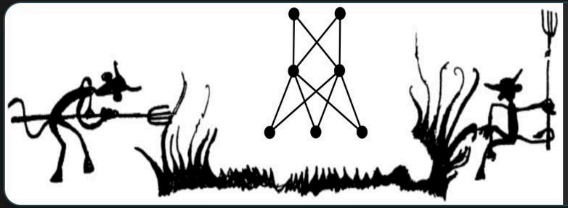
Shahverdi: Algebraic complexity and neurovariety of linear convolutional networks. **Acta Univ. Sapientiae Math.** 2025.

Usevich, Dérand, Borsoi, Clausel: Identifiability of Deep Polynomial Neural Networks. **NeurIPS Oral** 2025.

Auto

Neuroalgebraic Geometry

This page lists papers on the theory of algebraic neural networks.*



Algebra Unveils Deep Learning - An Invitation to Neuroalgebraic Geometry

Marchetti, Shahverdi, Mereta, Trager, Kohn — ICML Position Paper · 2025

Read paper

FILTERS

SORT BY: Activation | Architecture | Research Question | Year | Authors ↑

Showing 1–12 of 27 — Per page: 12 | 24 | 48 | 96

CNN · Linear · Learning Dynamics

The Riemannian Geometry associated to Gradient Flows of Linear Convolutional...

Achour, Kohn, Rauhut — AISTATS · 2026

Read paper

Polynomial · Optimization Degree

Algebraic Robustness Verification of Neural Networks

Alexandr, Duan, Montúfar — preprint · 2026

Read paper

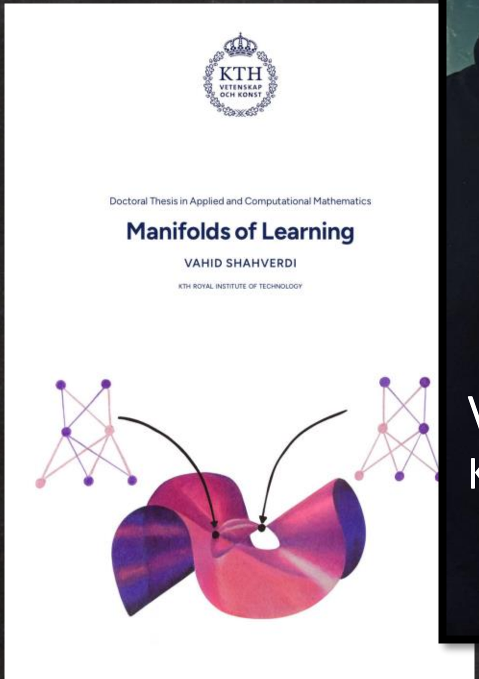
MLP · Monomial · Expressivity · Optimization Degree

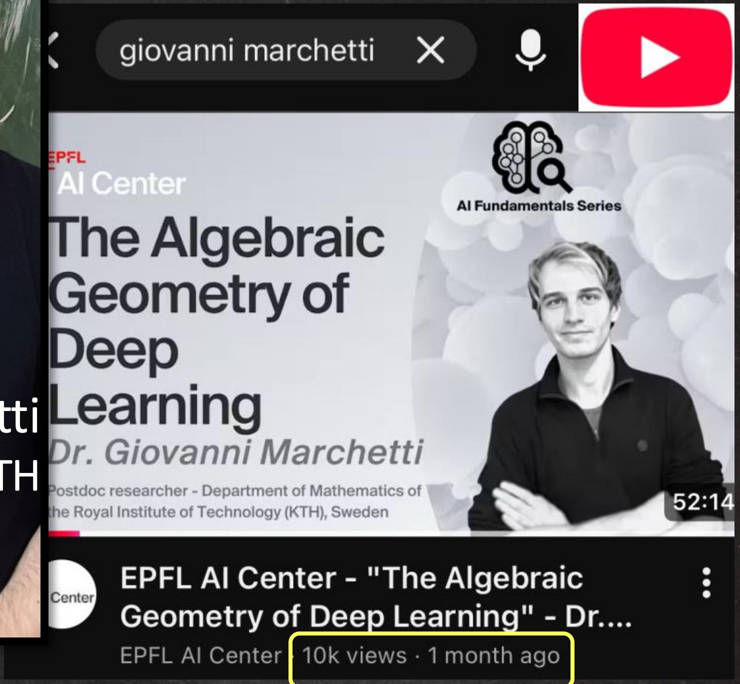
Geometry and Optimization of Shallow Polynomial Networks

Arjevani, Bruna, Kileel, Polak, Trager
SIAM Journal on Applied Algebra and Geometry · 2026

Read paper

special thanks to



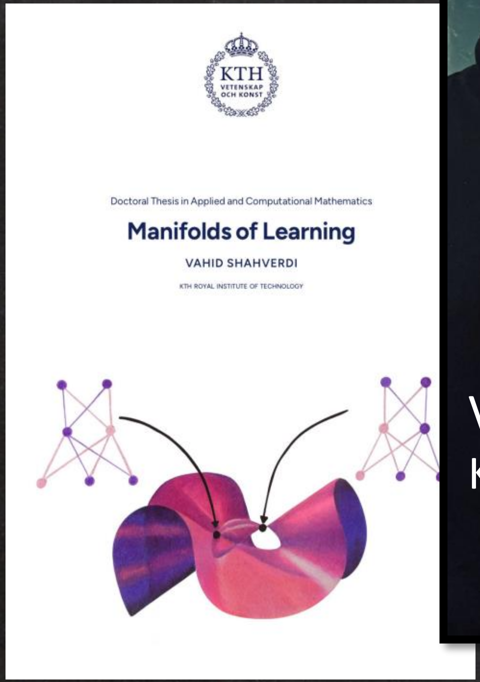


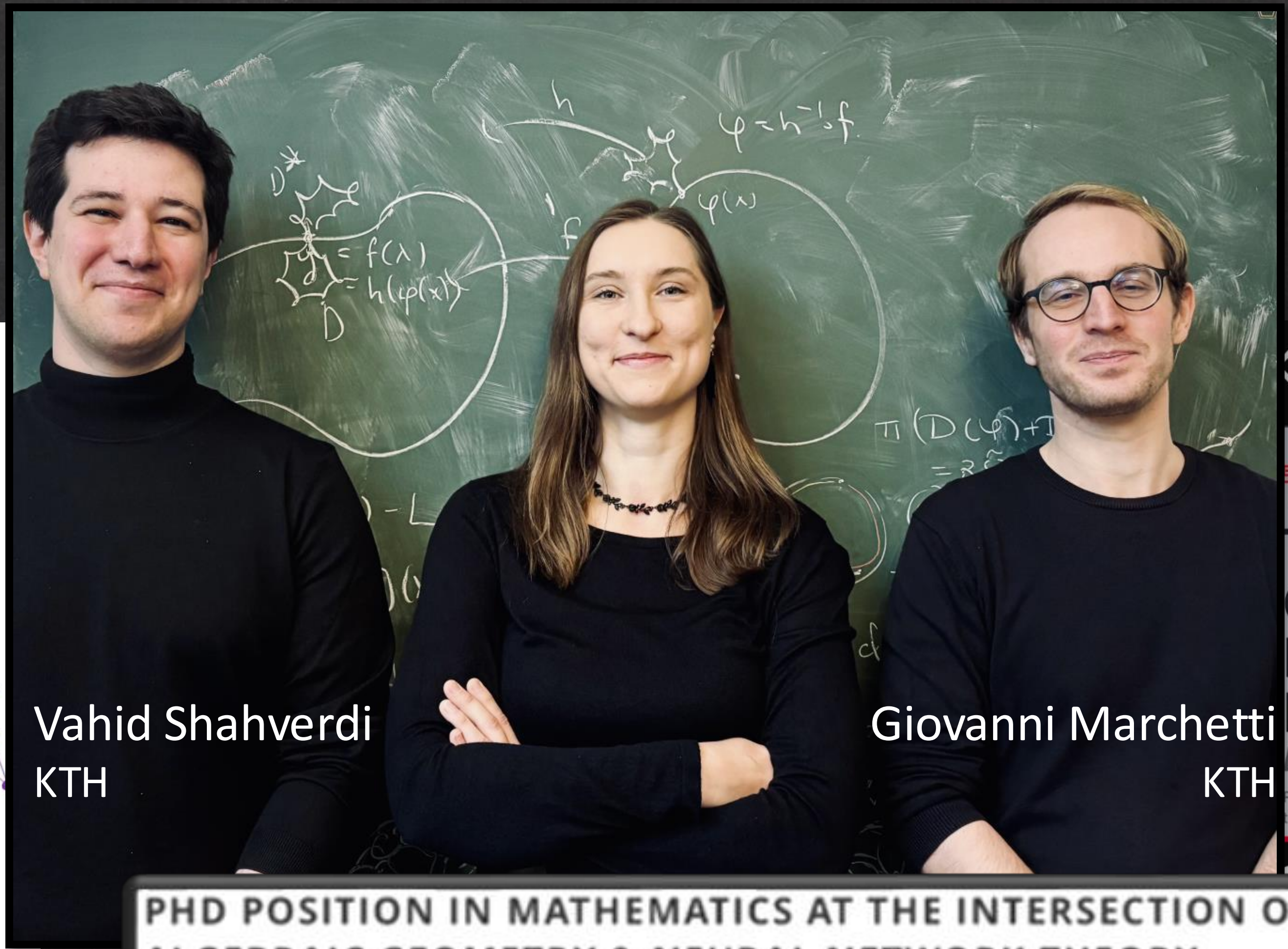
Matthew Trager
AWS AI Labs, NY

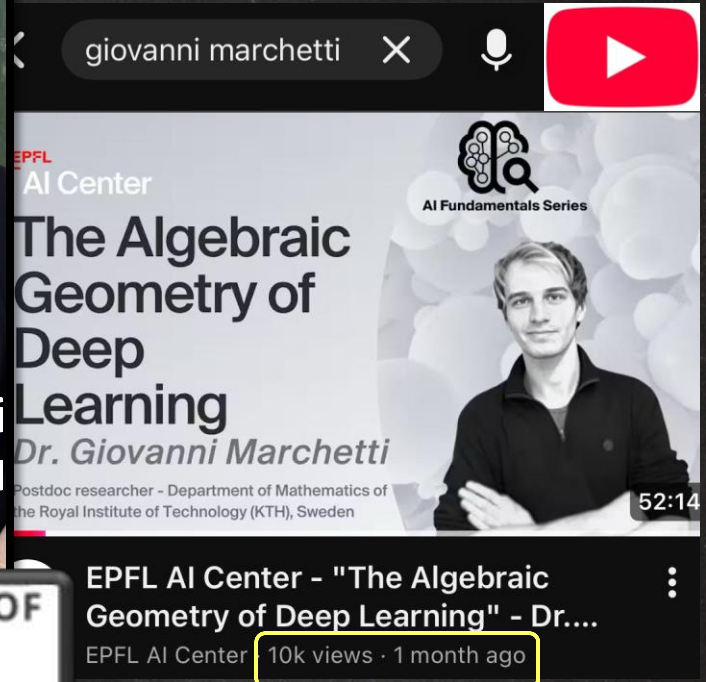
Joan Bruna
Courant Institute, NYU

Guido Montúfar
UCLA & MPI MiS Leipzig

special thanks to







PHD POSITION IN MATHEMATICS AT THE INTERSECTION OF ALGEBRAIC GEOMETRY & NEURAL NETWORK THEORY

KTH Royal Institute of Technology, Stockholm

Application deadline • May 20, 2026 11:59 PM CET

Matthew Trager
AWS AI Labs, NY

Joan Bruna
Courant Institute, NYU

Guido Montúfar
UCLA & MPI MiS Leipzig