

Neuromanifolds

The Geometry of Attention Networks and Polynomial Networks

Kathlén Kohn



based on joint works with

Joan Bruna

NYU

Nathan Henry

Univ. of Toronto

Giovanni Marchetti

KTH

Stefano Mereta

KTH

Guido Montúfar

UCLA, MPI MiS Leipzig

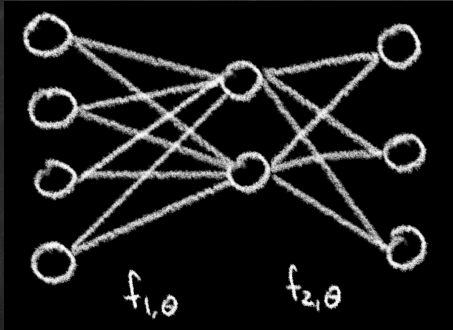
Vahid Shahverdi

KTH

Matthew Trager

Amazon

feedforward neural networks

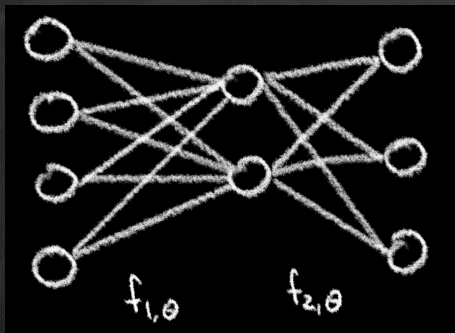


are parametrized families of functions

$$: \mathbb{R}^N \rightarrow \mathbb{R}^M;$$

$$\exists! f_L, \dots, f_1;$$

feedforward neural networks



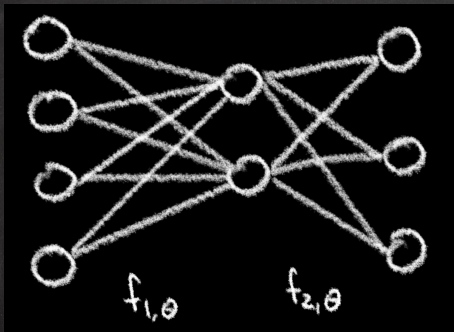
are parametrized families of functions

$$f_i : \mathbb{R}^N \rightarrow \mathbb{R}^M;$$

$$i = 1, \dots, L;$$

$$L = \# \text{ layers, } f_i = (w_{i,1}, \dots, w_{i,i}) \quad i = 1, \dots, L$$

feedforward neural networks



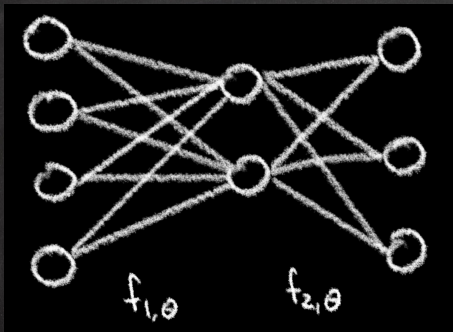
are parametrized families of functions

$$f_i : \mathbb{R}^N \rightarrow \mathbb{R}^M;$$

$$f = (f_L, \dots, f_1);$$

$L = \#$ layers, $f_i = (\dots \dots i)$ i ;
 $i : \mathbb{R} \rightarrow \mathbb{R}$ **activation**, i : affine linear

feedforward neural networks



$M = \text{im}(\) = \text{neuromanifold}$

it is a manifold with boundary and singularities

are parametrized families of functions

$$: \mathbb{R}^N \rightarrow M;$$

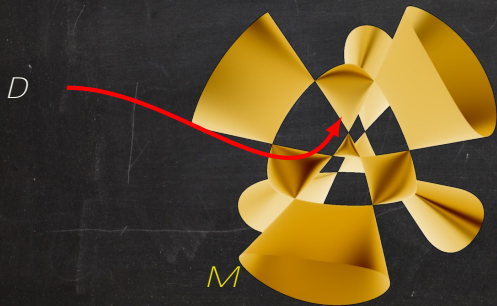
$$\exists! f_L, \dots, f_1;$$

$L = \#$ layers, $f_i = (\dots)$ i ;
 $i : \mathbb{R} \rightarrow \mathbb{R}$ **activation**, i affine linear

training a network

Given training data D , the goal is to minimize the **loss**

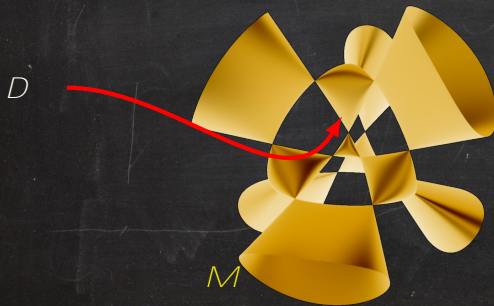
$$\mathbb{R}^N \ni M \in \mathbb{R}:$$



training a network

Given training data D , the goal is to minimize the **loss**

$$\mathbb{R}^N \ni M \in \mathcal{P}(\mathbb{R})$$



Geometric questions:

- ◆ How does the network architecture affect the geometry of the function space?
- ◆ How does the geometry of the function space impact the training of the network?

understanding networks via algebraic optimization

For piecewise algebraic activation, the neuromanifold is a **semi-algebraic set** (defined by polynomial equalities and inequalities).

understanding networks via algebraic optimization

For piecewise algebraic activation, the neuromanifold is a **semi-algebraic set** (defined by polynomial equalities and inequalities).

	activation	loss
Examples:	identity	
	ReLU	
	polynomial	

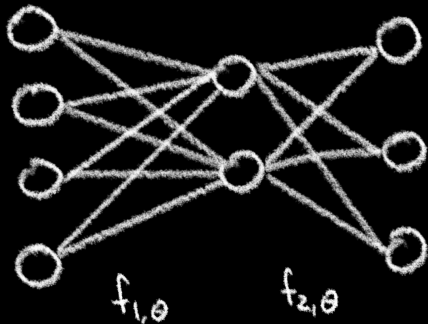
understanding networks via algebraic optimization

For piecewise algebraic activation, the neuromanifold is a **semi-algebraic set** (defined by polynomial equalities and inequalities).

	activation	loss	
Examples:	identity	squared-error loss	= Euclidean dist
	ReLU	Wasserstein distance	= polyhedral dist.
	polynomial	cross-entropy	= KL divergence

If the loss is also algebraic (or has at least algebraic derivatives), network training is an **algebraic optimization** problem.

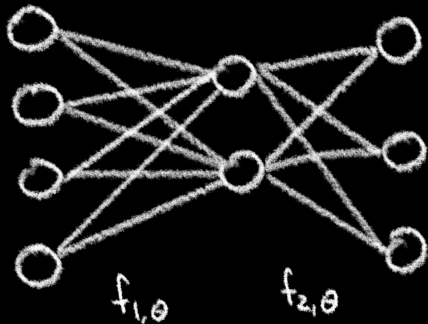
baby example: linear dense networks



In this example:

$$\begin{aligned} &: \mathbb{R}^{2 \times 4} \quad \mathbb{R}^{3 \times 2} \quad / \quad \mathbb{R}^{3 \times 4}; \\ &(W_1; W_2) \quad / \quad W_2 W_1: \end{aligned}$$

baby example: linear dense networks



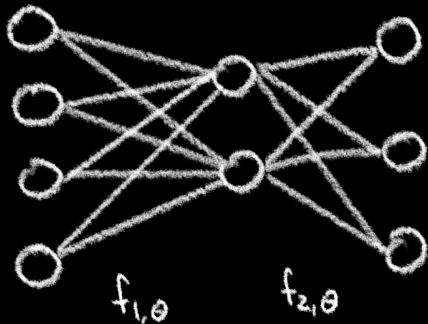
In this example:

$$: \mathbb{R}^{2 \times 4} \quad \mathbb{R}^{3 \times 2} \quad ; \quad \mathbb{R}^{3 \times 4};$$

$$(W_1; W_2) \quad ; \quad W_2 W_1:$$

$$M = fW \in \mathbb{R}^{3 \times 4} \quad ; \quad \text{rank}(W) \quad ; \quad 2g$$

baby example: linear dense networks



In this example:

$$\begin{aligned} &: \mathbb{R}^{2 \times 4} \quad \mathbb{R}^{3 \times 2} \quad ; \quad \mathbb{R}^{3 \times 4}; \\ &(W_1; W_2) \quad ; \quad W_2 W_1: \end{aligned}$$

$$M = fW \in \mathbb{R}^{3 \times 4} \quad \text{rank}(W) \quad 2g$$

In general:

$$\begin{aligned} &: \mathbb{R}^{k_1 \times k_0} \quad \mathbb{R}^{k_2 \times k_1} \quad \dots \quad \mathbb{R}^{k_L \times k_{L-1}} \quad ; \quad \mathbb{R}^{k_L \times k_0}; \\ &(W_1; W_2; \dots; W_L) \quad ; \quad W_L \quad W_2 W_1: \end{aligned}$$

$M = fW \in \mathbb{R}^{k_L \times k_0}$ $\text{rank}(W) = \min(k_0, \dots, k_L)g$ is an **algebraic variety** and we know its singularities etc.

example: attention networks

A single-layer lightning self-attention network with weights $Q, K \in \mathbb{R}^{a \times d}$ and $V \in \mathbb{R}^{d \times d}$ is

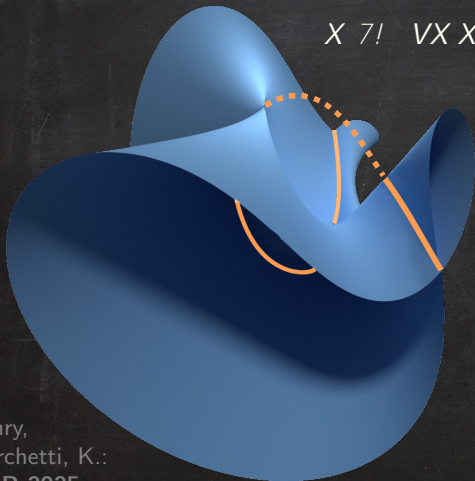
$$\mathbb{R}^{d \times t} \rightarrow \mathbb{R}^{d \times t};$$

$$X \mapsto VX X^T K^T QX:$$

example: attention networks

A single-layer lightning self-attention network with weights $Q, K \in \mathbb{R}^{a \times d}$ and $V \in \mathbb{R}^{d^0 \times d}$ is

$$\mathbb{R}^{d \times t} \rightarrow \mathbb{R}^{d^0 \times t};$$
$$X \mapsto VX X^{\top} K^{\top} QX;$$



A slice of the 5-dimensional neuromanifold \mathcal{M} for $a = d = t = 2; d^0 = 1$.

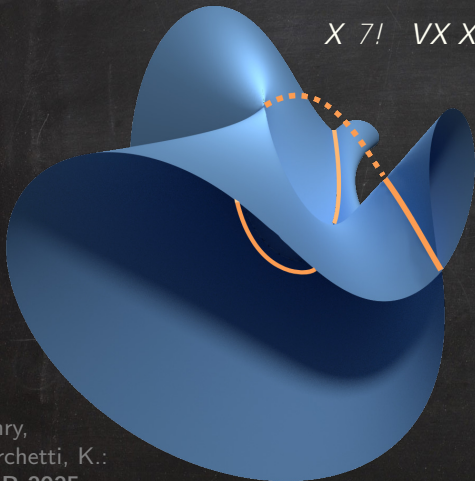
It is singular along the orange curve, and has boundary points where the curve leaves/enters \mathcal{M} .

example: attention networks

A single-layer lightning self-attention network with weights $Q, K \in \mathbb{R}^{a \times d}$ and $V \in \mathbb{R}^{d^0 \times d}$ is

$$\mathbb{R}^{d \times t} \rightarrow \mathbb{R}^{d^0 \times t};$$

$$X \mapsto VX X^T K^T QX;$$



A slice of the 5-dimensional neuromanifold \mathcal{M} for $a = d = t = 2; d^0 = 1$.

It is singular along the **orange curve**, and has boundary points where the **curve** leaves/enters \mathcal{M} .

It is not a variety, but a **semialgebraic set**.

a dictionary

machine learning

sample complexity & expressivity

subnetworks & implicit bias

identifiability & hidden symmetries

optimization & gradient descent

algebraic geometry

dimension, degree, covering number

singularities

fibers of the parametrization

critical point theory, discriminants,
dynamical invariants

dimension, degree, covering number

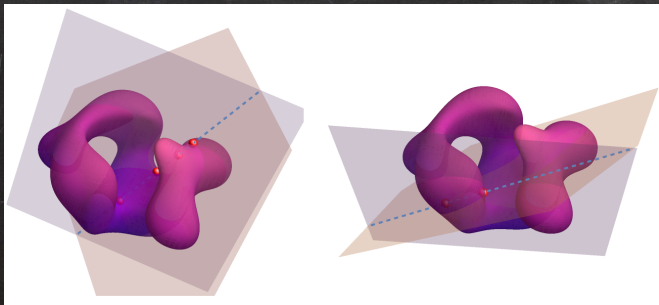
The **dimension** of the neuromanifold \mathcal{M} measures how many functions can be **exactly expressed** by the network.

dimension, degree, covering number

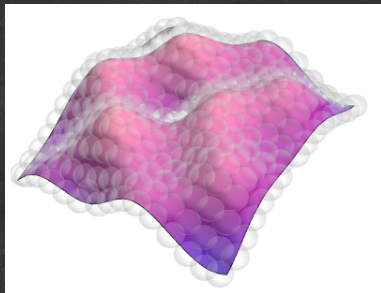
The **dimension** of the neuromanifold \mathcal{M} measures how many functions can be **exactly expressed** by the network.

The **degree** of an algebraic variety is the number of intersections (over \mathbb{C}) with a generic linear space (of the correct dimension).

It measures how curvy/twisted the variety is.

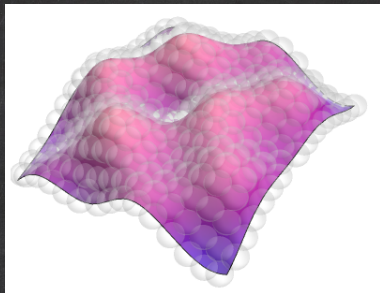


dimension, degree, covering number



covering number $N^\epsilon(M)$ = minimum number of metric balls of radius ϵ required to cover M

dimension, degree, covering number

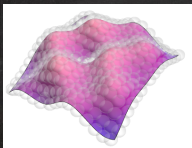


covering number $N''(\mathcal{M}) =$ minimum number of metric balls of radius ϵ required to cover \mathcal{M}

$$\log N''(\mathcal{M}) = O(\dim(\mathcal{M})) \log \frac{\text{degree}(\mathcal{M})}{\epsilon} + C$$

(cf. Weyl's Tube Formula)

dimension, degree, covering number

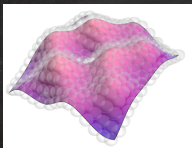


covering number $N_{\epsilon}(M)$ = minimum number of metric balls of radius ϵ required to cover M

relation to **sample complexity**:

the number of data samples required to infer the function that best approximates the distribution of data (with high probability, and within a given generalization loss margin ϵ) scales logarithmically in $N_{\epsilon}(M)$.

dimension, degree, covering number



covering number $N_\epsilon(M)$ = minimum number of metric balls of radius ϵ required to cover M

relation to **sample complexity**:

the number of data samples required to infer the function that best approximates the distribution of data (with high probability, and within a given generalization loss margin ϵ) scales logarithmically in $N_\epsilon(M)$.

relation to **approximative expressivity**:



the volume of the ϵ -tube around M measures how many functions can be approximated within an error of ϵ .

it is $N_\epsilon(M) \text{ vol}(\text{ball of radius } 2\epsilon)$

dimension, degree, covering number

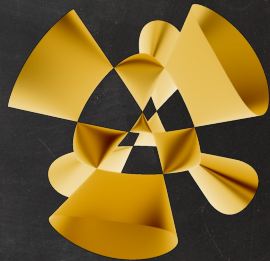
Takeaway

Dimension and degree are the most fundamental invariants of an algebraic neuromanifold.

They control metric quantities such as covering numbers, which in turn measure approximate expressivity and sample complexity.

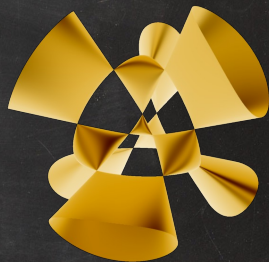
singularities

Singularities of a variety are points where the variety does not look locally like a smooth manifold.

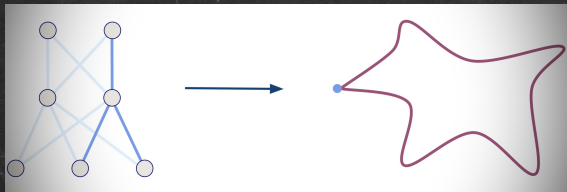


singularities

Singularities of a variety are points where the variety does not look locally like a smooth manifold.

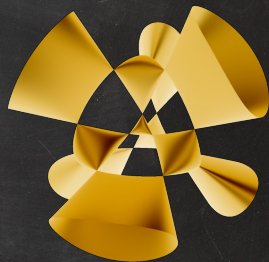


Conjecture: The **singularities** of neuromanifolds correspond to **subnetworks**.
(known for convolutional & fully-connected networks with polynomial activation)

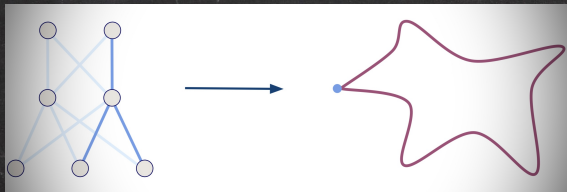


singularities

Singularities of a variety are points where the variety does not look locally like a smooth manifold.



Conjecture: The **singularities** of neuromanifolds correspond to **subnetworks**.
(known for convolutional & fully-connected networks with polynomial activation)



Potential explanation for *lottery ticket hypothesis*: the tendency of deep networks to discard weights during learning.

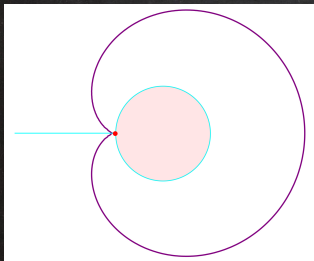
singularities

A **singularity** might, depending on its type, attract a large portion of the ambient space during training – explaining **implicit bias**.

singularities

A **singularity** might, depending on its type, attract a large portion of the ambient space during training – explaining **implicit bias**.

This is captured by the **Voronoi cell** of the singularity:



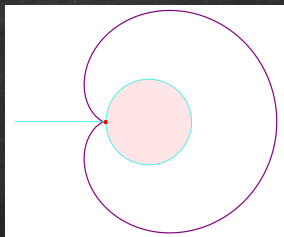
voronoi cells

Given a set $M \subset \mathbb{R}^n$, the **Voronoi cell** of $x \in M$ consists of all $u \in \mathbb{R}^n$ such that x is “closest” among all points in M .

M might be finite

voronoi cells

Given a set $M \subset \mathbb{R}^n$, the **Voronoi cell** of $x \in M$ consists of all $u \in \mathbb{R}^n$ such that x is “closest” among all points in M .



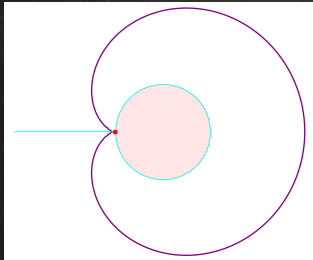
M might be finite

or a manifold, variety, semi-algebraic set, etc.

singularities

A **singularity** might, depending on its type, attract a large portion of the ambient space during training – explaining **implicit bias**.

This is captured by the **Voronoi cell** of the singularity:



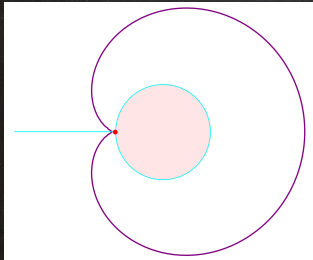
$M \subset \mathbb{R}^2$ is the purple curve

loss = Euclidean distance

singularities

A **singularity** might, depending on its type, attract a large portion of the ambient space during training – explaining **implicit bias**.

This is captured by the **Voronoi cell** of the singularity:



$M \subset \mathbb{R}^2$ is the purple curve

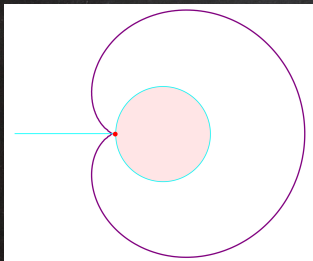
loss = Euclidean distance

at all smooth points $x \notin M$, the Voronoi cell is a **line segment**

singularities

A **singularity** might, depending on its type, attract a large portion of the ambient space during training – explaining **implicit bias**.

This is captured by the **Voronoi cell** of the singularity:



$M \subset \mathbb{R}^2$ is the purple curve

loss = Euclidean distance

at all smooth points $x \in M$, the Voronoi cell is a **line segment**

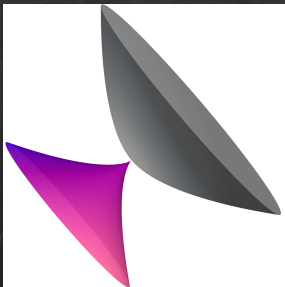
the Voronoi cell at the singularity is 2-dimensional, i.e., that point is the closest with **positive probability**

singularities

Takeaway

Singularities of the neuromanifold can introduce implicit biases in the learning process.

They often correspond to subnetworks, favoring the selection of simpler models.



bers of the parametrization

Recall: The neuromanifold is the image of parametrization map

$$\begin{aligned} &: \mathbb{R}^N \rightarrow M; \\ & \quad \theta \mapsto f_L; \quad \dots \quad f_1; \end{aligned}$$

Identifiability / hidden symmetries:

Which network parameters give rise to the same function?

fibers of the parametrization

Recall: The neuromanifold is the image of parametrization map

$$\begin{aligned} &: \mathbb{R}^N \rightarrow M; \\ & \theta \mapsto f_L; \dots; f_1; \dots \end{aligned}$$

Identifiability / hidden symmetries:

Which network parameters give rise to the same function?

In algebraic geometry terms:

Given $f \in M$, which parameters are in the **fiber** $f^{-1}(f)$?

fibers of the parametrization

Recall: The neuromanifold is the image of parametrization map

$$f : \mathbb{R}^N \rightarrow M;$$
$$f = (f_1, \dots, f_L);$$

Identifiability / hidden symmetries:

Which network parameters give rise to the same function?

In algebraic geometry terms:

Given $f \in M$, which parameters are in the **fiber** $f^{-1}(f)$?

Fiber-dimension theorem:

The dimension of the image of an algebraic map equals the co-dimension of its generic fiber.
(nonlinear version of rank-nullity theorem)

Parameters of the parametrization

Recall: The neuromanifold is the image of parametrization map

$$f: \mathbb{R}^N \rightarrow M;$$
$$f = (f_1, \dots, f_L);$$

More generally: All geometric features of the neuromanifold are caused by

fibers of the parametrization

Recall: The neuromanifold is the image of parametrization map

$$f: \mathbb{R}^N \rightarrow M;$$
$$f = (f_1, \dots, f_L);$$

More generally: All geometric features of the neuromanifold are caused by

For instance, singularities on M can arise in 2 ways:

- ◆ from critical points of
- ◆ from special (i.e., non-generic) fibers of

example: polynomial convolutional networks

We now consider convolutional networks

where the activation function is a monomial: $\sigma(x) = x^r$.

example: polynomial convolutional networks

We now consider convolutional networks

where the activation function is a monomial: $(x) = x^r$.

Weierstrass Approximation Theorem:

Any activation function can be approximated by polynomial ones.

Any CNN neuromanifold can be approximated by polynomial ones.

example: polynomial convolutional networks

$$(x) = x^r$$

Theorem: Let $r > 1$.

The neuromanifold is an **algebraic variety** (i.e., described by polynomial equations) and closed in Euclidean topology.

example: polynomial convolutional networks

$$(x) = x^r$$

Theorem: Let $r > 1$.

The neuromanifold is an **algebraic variety** (i.e., described by polynomial equations) and closed in Euclidean topology.

For a generic function $f \in M$, the only **symmetries** in the **fiber** $\pi^{-1}(f)$ are rescalings of the layers.

example: polynomial convolutional networks

$$(x) = x^r$$

Theorem: Let $r > 1$.

The neuromanifold is an **algebraic variety** (i.e., described by polynomial equations) and closed in Euclidean topology.

For a generic function $f \in M$, the only **symmetries** in the **fiber** $\pi^{-1}(f)$ are rescalings of the layers.

After modding out the layer scaling, the network parametrization map becomes

- ◆ an isomorphism almost everywhere

example: polynomial convolutional networks

$$(x) = x^r$$

Theorem: Let $r > 1$.

The neuromanifold is an **algebraic variety** (i.e., described by polynomial equations) and closed in Euclidean topology.

For a generic function $f \in M$, the only **symmetries** in the **fiber** $\pi^{-1}(f)$ are rescalings of the layers.

After modding out the layer scaling, the network parametrization map becomes

- ◆ an isomorphism almost everywhere
- ◆ that has finite fibers (, **singularities**)
- ◆ and is regular (constant-rank Jacobian)

example: polynomial convolutional networks

$$(x) = x^r$$

Theorem: Let $r > 1$.

The neuromanifold is an **algebraic variety** (i.e., described by polynomial equations) and closed in Euclidean topology.

For a generic function $f \in M$, the only **symmetries** in the **fiber** $\pi^{-1}(f)$ are rescalings of the layers.

After modding out the layer scaling, the network parametrization map becomes

- ◆ an isomorphism almost everywhere
- ◆ that has finite fibers (, **singularities**)
- ◆ and is regular (constant-rank Jacobian)

The **singularities** correspond to **subnetworks**.

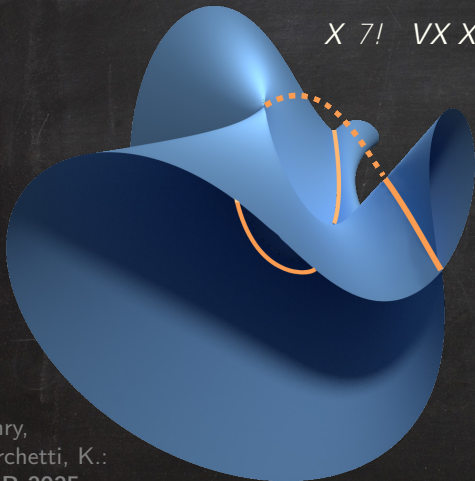
Shahverdi, Marchetti, K.:

comparison: lightning self-attention

A single-layer lightning self-attention network with weights $Q, K \in \mathbb{R}^{a \times d}$ and $V \in \mathbb{R}^{d_0 \times d}$ is

$$\mathbb{R}^{d \times t} \rightarrow \mathbb{R}^{d_0 \times t};$$

$$X \mapsto VX X^T K^T QX;$$

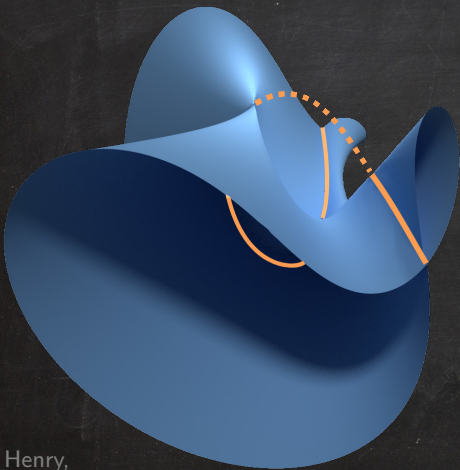


The neuromanifold is semialgebraic but not a variety (polynomial inequalities needed!)

It has both nodal and cuspidal singularities.

comparison: lightning self-attention

$$VXX > K > QX$$

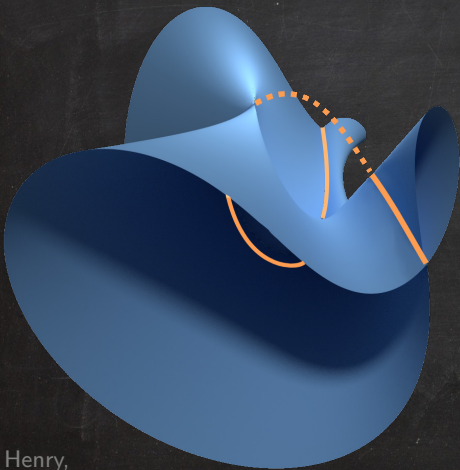


cusps

- , boundary points
- , Jacobian rank drops

comparison: lightning self-attention

$$VXX > K > QX$$



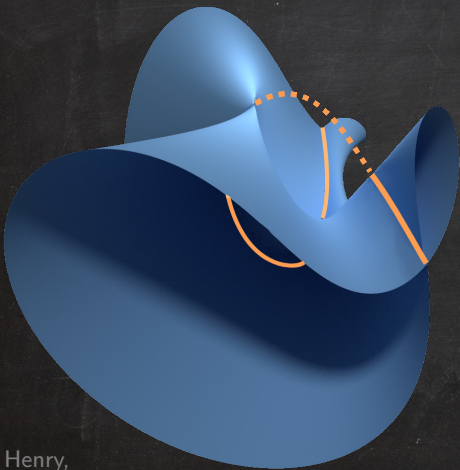
cusps

- , boundary points
- , Jacobian rank drops

Theorem: For generic $f \in M$, the only **symmetries** in the **fiber** $f^{-1}(f)$ are the “obvious” ones:

comparison: lightning self-attention

$$VXX > K > QX$$



cusps

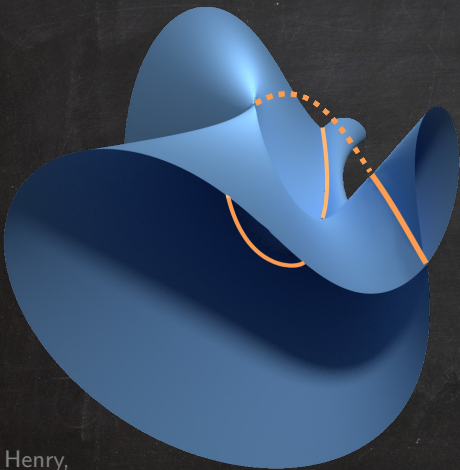
- , boundary points
- , Jacobian rank drops

Theorem: For generic $f \in M$,
the only **symmetries** in the **fiber**
 $f^{-1}(f)$ are the “obvious” ones:

- ◆ layer rescalings

comparison: lightning self-attention

$$VXX > K > QX$$



cusps

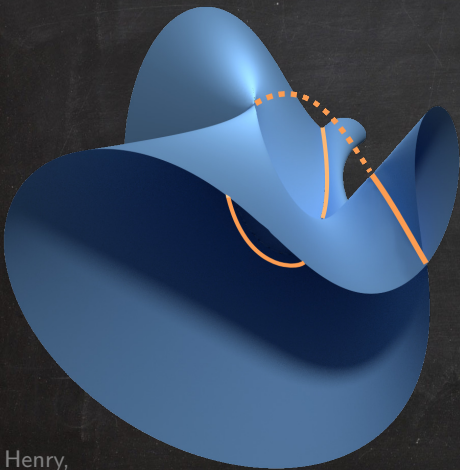
- , boundary points
- , Jacobian rank drops

Theorem: For generic $f \in M$, the only **symmetries** in the **fiber** $\pi^{-1}(f)$ are the “obvious” ones:

- ◆ layer rescalings
- ◆ $GL(a)$ -symmetries of K and Q in each layer

comparison: lightning self-attention

$$VXX > K > QX$$



cusps

- , boundary points
- , Jacobian rank drops

Theorem: For generic $f \in M$, the only **symmetries** in the **fiber** $\pi^{-1}(f)$ are the “obvious” ones:

- ◆ layer rescalings
- ◆ $GL(a)$ -symmetries of K and Q in each layer
- ◆ $GL(d)$ -symmetries of V and $K > Q$ of neighboring layers

bers of the parametrization

Takeaway

Fibers of the parameterization control the dimension and symmetries of the neuromanifold.

Together with the parameterization's critical points, they explain the singularities of the neuromanifold.

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow \mathbb{R}$$

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow M \rightarrow \mathbb{R}$$

Critical points of L_D arise in various ways:

1. they can be caused by the parametrization
(i.e., $\exists \text{Crit}(\gamma)$ such that $\gamma \in \text{Crit}(L_D)$ but $(\gamma) \notin \text{Crit}(L_D)$)

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow M \rightarrow \mathbb{R}$$

Critical points of L_D arise in various ways:

1. they can be caused by the parametrization
(i.e., $\exists \text{Crit}(\gamma)$ such that $\gamma \in \text{Crit}(L_D)$ but $(\gamma) \notin \text{Crit}(L_D)$)
spurious critical points

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow \mathbb{R}$$

Critical points of L_D arise in various ways:

1. they can be caused by the parametrization
(i.e., $\exists \text{Crit}(\cdot)$ such that $\exists \text{Crit}(L_D)$ but $(\cdot) \notin \text{Crit}(L_D)$)
spurious critical points

e.g. appear as local minima in polynomial MLPs with positive probability
but not in polynomial CNNs

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow \mathbb{R}$$

Critical points of L_D arise in various ways:

1. they can be caused by the parametrization
(i.e., $\theta \in \text{Crit}(\phi)$ such that $\theta \in \text{Crit}(L_D)$ but $\theta \notin \text{Crit}(\phi)$)
spurious critical points
e.g. appear as local minima in polynomial MLPs with positive probability
but not in polynomial CNNs
2. they correspond to critical points of the loss in function space
(i.e., $\theta \in \text{Crit}(L_D)$ and $\theta \in \text{Crit}(\phi)$)

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow M \xrightarrow{P} \mathbb{R}$$

Critical points of L_D arise in various ways:

1. they can be caused by the parametrization
(i.e., $\exists \text{Crit}(\cdot)$ such that $\exists \text{Crit}(L_D)$ but $(\cdot) \notin \text{Crit}(\cdot_D)$)
spurious critical points

e.g. appear as local minima in polynomial MLPs with positive probability
but not in polynomial CNNs

2. they correspond to critical points of the loss in function space
(i.e., $\exists \text{Crit}(L_D)$ and $(\cdot) \in \text{Crit}(\cdot_D)$)
the function (\cdot) can be either a
 - a) singular point on M or
 - b) in the smooth locus of M

critical point theory & discriminants

Goal: minimize the loss

$$L_D: \mathbb{R}^N \rightarrow M \xrightarrow{p} \mathbb{R}$$

Critical points of L_D arise in various ways:

1. they can be caused by the parametrization
(i.e., $\exists \text{Crit}(\cdot)$ such that $\exists \text{Crit}(L_D)$ but $(\cdot) \notin \text{Crit}(\cdot)$)
spurious critical points

e.g. appear as local minima in polynomial MLPs with positive probability
but not in polynomial CNNs

2. they correspond to critical points of the loss in function space
(i.e., $\exists \text{Crit}(L_D)$ and $(\cdot) \in \text{Crit}(\cdot)$)
the function (\cdot) can be either a
 - a) singular point on M or
 - b) in the smooth locus of M

Morse theory

critical point theory & discriminants

for algebraic optimization problems (e.g. mean squared error or cross entropy loss), the number of complex critical points of L_D is constant for generic D

critical point theory & discriminants

for algebraic optimization problems (e.g. mean squared error or cross entropy loss), the number of complex critical points of L_D is constant for generic D
measures intrinsic **optimization degree**

over \mathbb{R} , the number or **type** (local / global minima, strict / non-strict saddle, etc.) of the critical points changes when D crosses an algebraic **discriminant hypersurface**

critical point theory & discriminants

for algebraic optimization problems (e.g. mean squared error or cross entropy loss), the number of complex critical points of L_D is constant for generic D
measures intrinsic **optimization degree**

over \mathbb{R} , the number or **type** (local / global minima, strict / non-strict saddle, etc.) of the critical points changes when D crosses an algebraic **discriminant hypersurface**

over \mathbb{C} : always 4
critical points

over \mathbb{R} : 4 or 2 critical
points

discriminant = dashed

critical point theory, discriminants, dynamical invariants

Takeaway

The critical points of the loss arise from the geometry of the neuromanifold and its parametrization.

Their number and type can change suddenly as data crosses discriminants.

Moreover, algebraic invariants of gradient flow govern the training dynamics...

many future questions

- ◆ Describe all **singularities** of attention neuromanifolds explicitly, and compute their Voronoi cells. (**implicit bias?**)
- ◆ Compare the type of critical points and more generally the loss landscape of
 - ◆ attention networks
 - ◆ polynomial convolutional networks
 - ◆ polynomial dense networks
- ◆ How do skip connections and inhomogeneous activations regularize μ (i.e., less spurious critical points) and smoothen out singularities?
- ◆ What happens to the neuromanifold when imposing group equivariance?
- ◆ What about ReLU networks, or more generally piecewise rational activation?
- ◆ Beyond algebraic geometry: tame geometry of o-minimal structures

thanks for your attention!

machine learning

algebraic geometry

sample complexity & expressivity

dimension, degree, covering number

subnetworks & implicit bias

singularities

identifiability & hidden symmetries

fibers of the parametrization

optimization & gradient descent

critical point theory, discriminants,
dynamical invariants

An Invitation to Neuroalgebraic Geometry

Giovanni Luca Marchetti^{*1} Vahid Shahverdi^{*1} Stefano Mereta^{*1} Matthew Trager^{*2} Kathlén Kohn^{*1}

Abstract

In this expository work, 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 activations, whose associated function spaces are semi-algebraic varieties. We outline a dictionary between algebro-geometric invariants of these varieties, such as dimension, degree, and singularities, and fundamental aspects of machine learning, such as sample complexity, expressivity, training dynamics, and implicit bias.

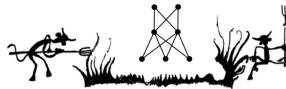
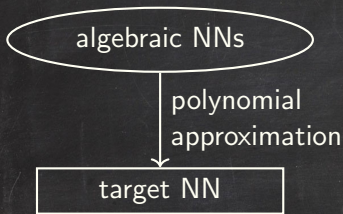


Figure 1. A neural variation of a celebrated doodle from the algebraic geometry literature (Grothendieck, 1968).

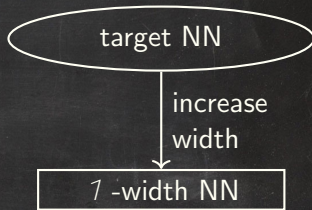
Neuroalgebraic Geometry



studies **nonlinear** models
in **finite-dimensional** ambient space

aims to draw conclusions
in the limit

NTK approach



studies **linearized** models
in **1-dimensional** ambient space

aims to draw conclusions
from the limit



Department of Mathematics and
Mathematical Statistics

Geometric Deep Learning

August 2025
Umeå

invited speakers:

Elisenda Grigsby
Boston College

Emanuele Rodolá
University of Rome

emphasis:
presentations by young
researchers

August 19-21

Axel Flinth (Umeå Univ), Jan E. Gerken (Chalmers), Kathlén Kohn (KTH),
Giovanni Marchetti (KTH), Stefano Mereta (KTH), Fredrik Ohlsson (Umeå Univ)