

Algebraic Neural Network Theory



Kathlén Kohn

KTH
Digital Futures

digital futures

WASP | WALLENBERG AI,
AUTONOMOUS SYSTEMS
AND SOFTWARE PROGRAM

**SF
SG** SWEDISH
FOUNDATIONS'
STARTING GRANT



Ragnar Söderbergs
STIFTELSE



Göran Gustafssons Stiftelser

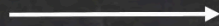
deep learning in a nutshell



parameter space

θ

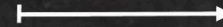
learnable weights



function space

f_{θ}

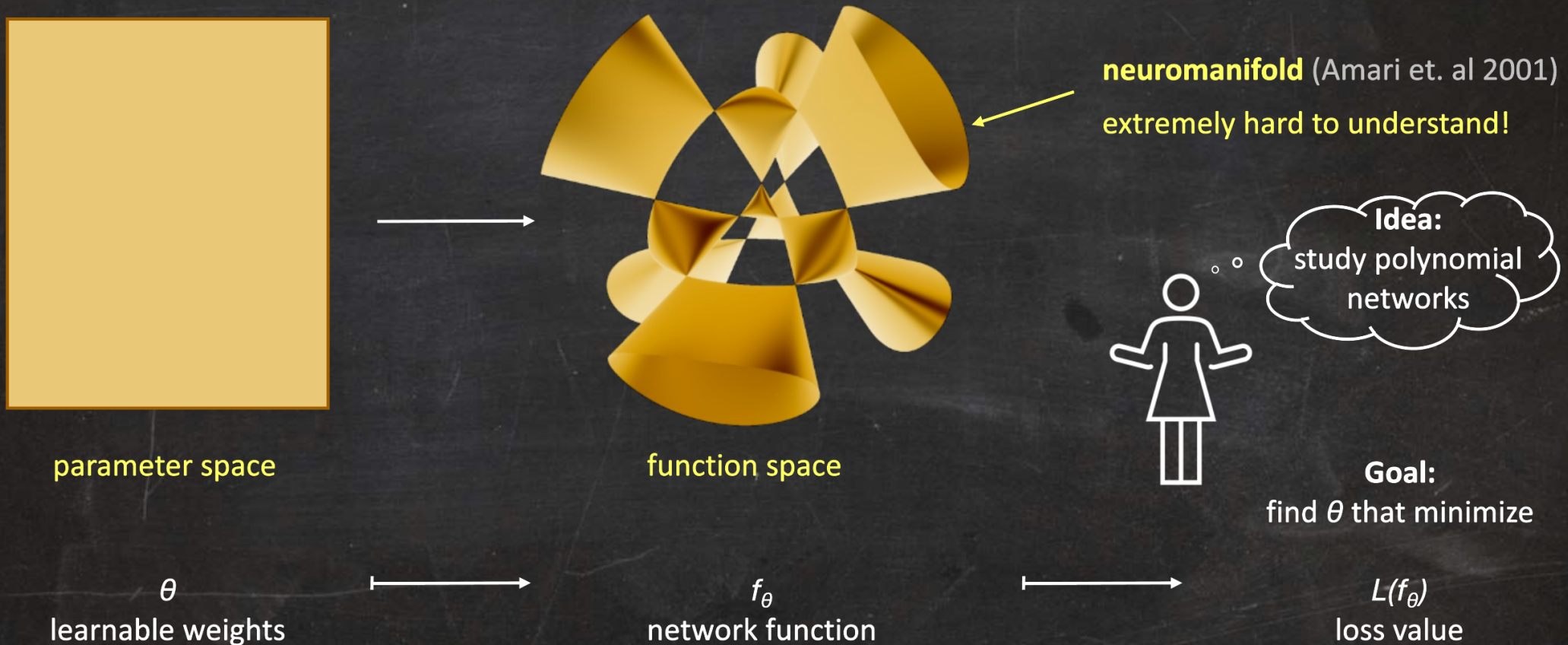
network function



Goal:
find θ that minimize

$L(f_{\theta})$
loss value

deep learning in a nutshell



Example: MLPs multilayer perceptrons

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

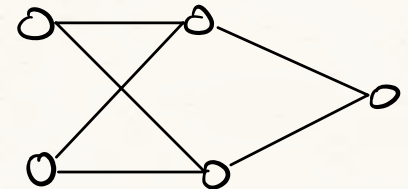
$\alpha_i =$ learnable affine linear functions

$\sigma =$ nonlinear activation function, applied entrywise

we assume: σ is a univariate **polynomial**

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

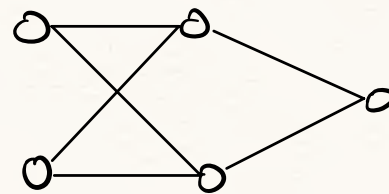
$$[e \ f] \sigma \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$



Which functions does this MLP parametrize?

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

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



Which functions does this MLP parametrize?

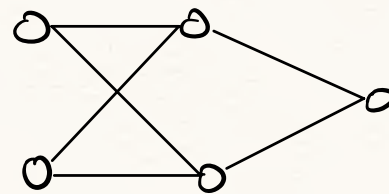
$$\begin{aligned} & e(ax+by)^2 + f(cx+dy)^2 \\ &= \underbrace{(a^2e + c^2f)}_A x^2 + \underbrace{2(ab e + cd f)}_B xy + \underbrace{(b^2e + d^2f)}_C y^2 \end{aligned}$$

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

← homogeneous quadratic polynomials in x,y
i.e., are all values for A, B, C possible?

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

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



Which functions does this MLP parametrize?

$$\begin{aligned} & e(ax+by)^2 + f(cx+dy)^2 \\ &= \underbrace{(a^2e + c^2f)}_A x^2 + \underbrace{2(ab e + cd f)}_B xy + \underbrace{(b^2e + d^2f)}_C y^2 \end{aligned}$$

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

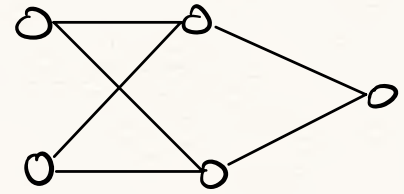
← homogeneous quadratic polynomials in x,y
i.e., are all values for A, B, C possible?

YES

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

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

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



Which functions does this MLP parametrize?

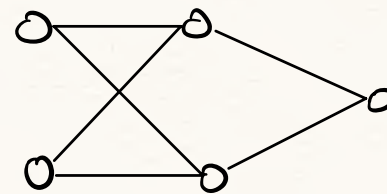
$$\begin{aligned} & e(ax+by)^3 + f(cx+dy)^3 \\ &= \underbrace{(a^3e + c^3f)}_A x^3 + \underbrace{3(a^2be + c^2df)}_B x^2y + \underbrace{3(ab^2e + cd^2f)}_C xy^2 + \underbrace{(b^3e + d^3f)}_D y^3 \end{aligned}$$

Can you obtain all 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?

$$\begin{aligned} & e(ax+by)^3 + f(cx+dy)^3 \\ &= \underbrace{(a^3e + c^3f)}_A x^3 + \underbrace{3(a^2be + c^2df)}_B x^2y + \underbrace{3(ab^2e + cd^2f)}_C xy^2 + \underbrace{(b^3e + d^3f)}_D y^3 \end{aligned}$$

Can you obtain all 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. $A = 1$
 $B = 0$
 $C = -1$
 $D = 0$

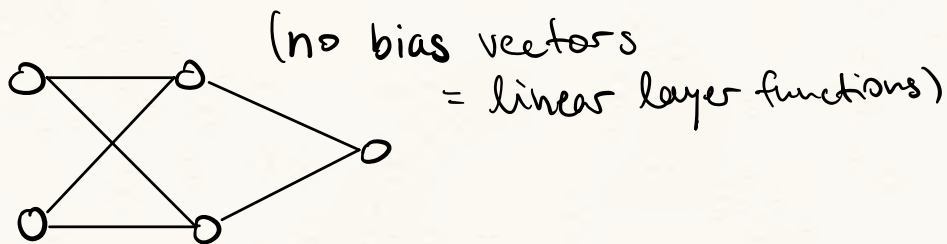


Neuromanifolds

A **parametric machine learning** model is a map $\mu: \Theta \times X \rightarrow Y$.
 parameters $\uparrow$ $\uparrow$ $\uparrow$
 Θ X Y
 inputs outputs

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

**Example
MLPs:**



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

$$\Rightarrow \mathcal{M} = \mathbb{R}[x, y]_2$$

$$\sigma(x) = x^3$$

$$\Rightarrow \mathcal{M} \subsetneq \mathbb{R}[x, y]_3$$

$$\sigma(x) = x$$

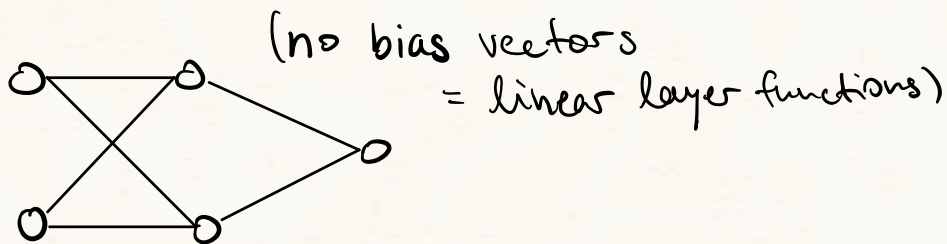
$$\Rightarrow ?$$

Neuromanifolds

A **parametric machine learning** model is a map $\mu: \Theta \times X \rightarrow Y$.
 parameters $\uparrow$ $\uparrow$ $\uparrow$
 Θ X Y
 inputs outputs

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

**Example
MLPs:**



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

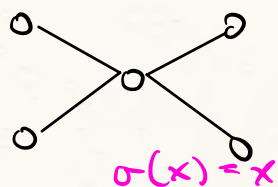
$$\Rightarrow \mathcal{M} = \mathbb{R}[x, y]_2$$

$$\sigma(x) = x^3$$

$$\Rightarrow \mathcal{M} \subsetneq \mathbb{R}[x, y]_3$$

$$\sigma(x) = x$$

$$\Rightarrow \mathcal{M} = \mathbb{R}^{1 \times 2}$$



$$\begin{bmatrix} c \\ a \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

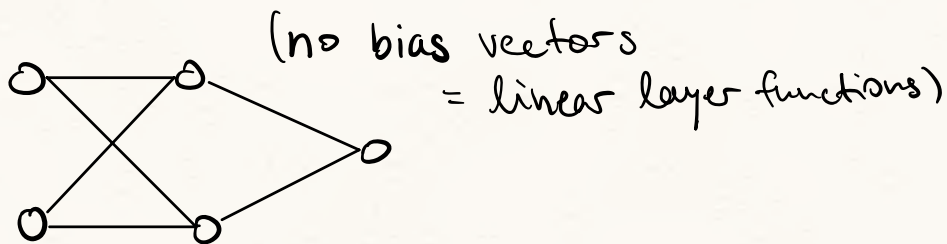
$$\Rightarrow \mathcal{M} = ?$$

Neuromanifolds

A **parametric machine learning** model is a map $\mu: \Theta \times X \rightarrow Y$.
 parameters $\uparrow$ $\uparrow$ $\uparrow$
 Θ X Y
 inputs outputs

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

**Example
MLPs:**



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

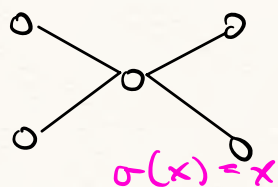
$$\Rightarrow \mathcal{M} = \mathbb{R}[x, y]_2$$

$$\sigma(x) = x^3$$

$$\Rightarrow \mathcal{M} \subsetneq \mathbb{R}[x, y]_3$$

$$\sigma(x) = x$$

$$\Rightarrow \mathcal{M} = \mathbb{R}^{1 \times 2}$$



$$\begin{bmatrix} c \\ a \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Rightarrow \mathcal{M} = \{ W \in \mathbb{R}^{2 \times 2} \mid \text{rk}(W) \leq 1 \}$$

Linear MLPs: $\alpha_L \circ \dots \circ \alpha_2 \circ \alpha_1$, where

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

$$\Rightarrow \mathcal{M} = \{W \in \mathbb{R}^{d_L \times d_0} \mid \text{rk}(W) \leq \min\{d_0, d_1, \dots, d_L\}\}$$

Linear MLPs: $\alpha_L \circ \dots \circ \alpha_2 \circ \alpha_1$, where

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

$$\Rightarrow \mathcal{M} = \{W \in \mathbb{R}^{d_L \times d_0} \mid \text{rk}(W) \leq \min\{d_0, d_1, \dots, d_L\}\}$$

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

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

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

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

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

Linear MLPs: $\alpha_L \circ \dots \circ \alpha_2 \circ \alpha_1$, where
 $\alpha_i: \mathbb{R}^{d_{i-1}} \rightarrow \mathbb{R}^{d_i}$ linear

$$\Rightarrow \mathcal{M} = \{W \in \mathbb{R}^{d_L \times d_0} \mid \text{rk}(W) \leq \min\{d_0, d_1, \dots, d_L\}\}$$

Polynomial MLPs: $\alpha_L \circ \sigma \circ \dots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$, where
 $\alpha_i: \mathbb{R}^{d_{i-1}} \rightarrow \mathbb{R}^{d_i}$ affine linear

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

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

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

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

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

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:

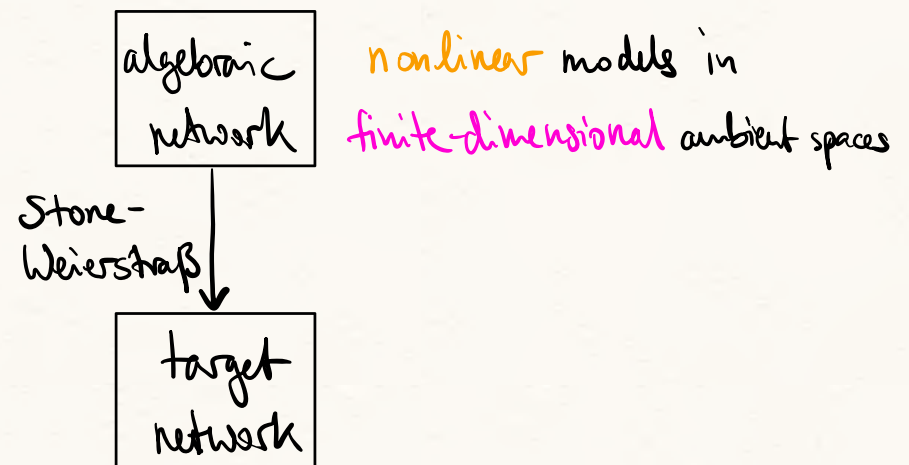
Let $\sigma \in M$. Set

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

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

polynomials are the choice
to approximate networks with
finite-dimensional models

AG approach



neural
tangent
kernel
NTK approach

target
network

increase
width

∞ -width
network

linearized models
of ∞ dimension

algebraic
geometry
AG approach

algebraic
network

Stone-
Weierstraß

target
network

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)

> 4000 citations

Network training = 'distance' minimization

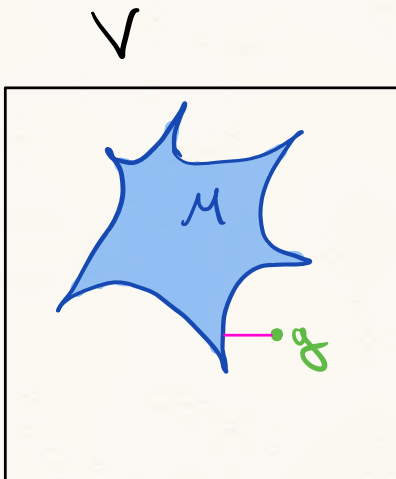
Let $\mathcal{M} \subseteq V := \left(\mathbb{R}[x_1, \dots, x_{d_0}] \leq \mathbb{D} \right)^{d_L}$,
 $\nwarrow$ neuromanifold

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

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

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

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



Why?

Network training = 'distance' minimization

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

$\nwarrow$ neuromanifold

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

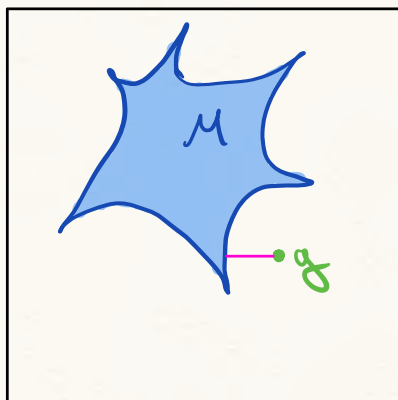
$\nwarrow$ mean squared error

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

$\nwarrow$ [dist(f,g) = 0 possible for f ≠ g]

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

V



Assume: $d_L = 1$

Let $v_D: (x_1, \dots, x_{d_0}) \mapsto (\text{all monomials in } x_1, \dots, x_{d_0} \text{ of degree } \leq \mathbb{D})$,
 c_f be coefficient vector of $f \in V$ such that $f(x) = v_D(x) \cdot c_f$,

Veronese embedding ↘

Network training = 'distance' minimization

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

$\nwarrow$ neuromanifold

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

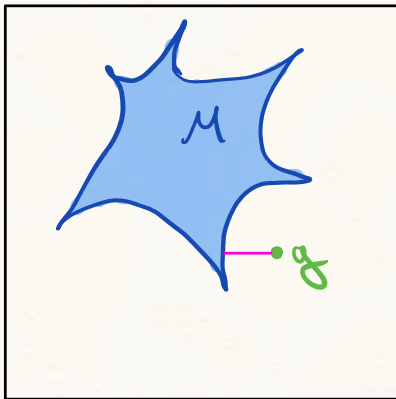
$\nwarrow$ mean squared error

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

$\nwarrow$ [dist(f,g) = 0 possible for $f \neq g$]

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

V



Assume: $d_L = 1$

Let $v_D: (x_1, \dots, x_{d_0}) \mapsto$ (all monomials in $x_1, \dots, x_{d_0}$ of degree $\leq \mathbb{D}$),

c_f be coefficient vector of $f \in V$ such that $f(x) = v_D(x) \cdot c_f$,

A & B matrices whose rows are $v_D(a)$ & b , resp., over all $(a,b) \in S$

$$\Rightarrow \mathcal{L}(f) = \|A c_f - B\|^2$$

Veronese embedding ↘

Network training = 'distance' minimization

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

$\nwarrow$ *neuromanifold*

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

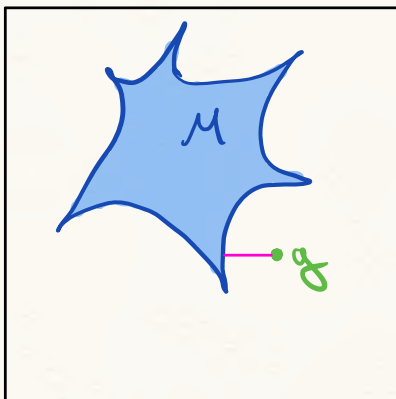
$\nwarrow$ *mean squared error*

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

$\nwarrow$ [dist(f,g) = 0 possible for $f \neq g$]

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

V



Assume: $d_L = 1$

Let $v_D: (x_1, \dots, x_{d_0}) \mapsto (\text{all monomials in } x_1, \dots, x_{d_0} \text{ of degree } \leq \mathbb{D})$,

c_f be coefficient vector of $f \in V$ such that $f(x) = v_D(x) \cdot c_f$,

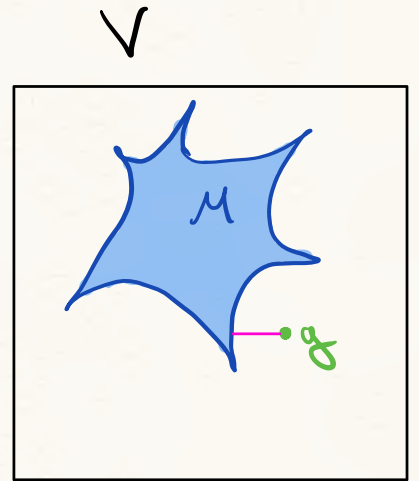
A & B matrices whose rows are $v_D(a)$ & b , resp., over all $(a,b) \in S$

Veronese embedding $\curvearrowright$

$$\Rightarrow \mathcal{L}(f) = \|A c_f - B\|^2 = \|c_f - A^+ B\|^2 + \text{const.}$$

$\nwarrow$ *pseudoinverse*
 $\nwarrow$ $\|c\|_Q := c^T Q c$

$$\arg\min_{f \in \mathcal{M}} L(f) = \arg\min_{f \in \mathcal{M}} \|C_f - A^+ B\|_{A^T A}^2$$



Observations ($d_L=1$):

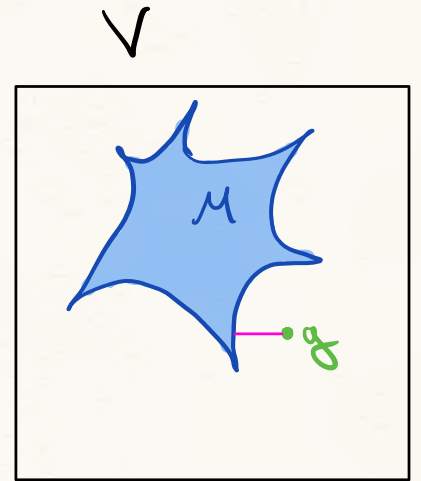
① $A^T A$ depends only on input data,
 $A^+ B$ on both input & output

② $A^T A \in \mathbb{R}^{\dim V \times \dim V}$ is rank-deficient whenever $|S| < \dim V \rightarrow$ pseudometric

③

(LLMs: $|S| < \dim M$)

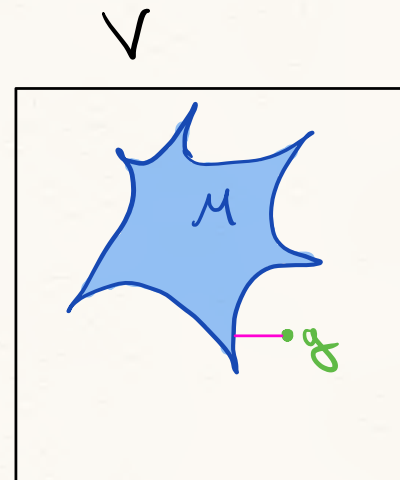
$$\arg\min_{f \in \mathcal{M}} L(f) = \arg\min_{f \in \mathcal{M}} \|C_f - A^+ B\|_{A^T A}^2$$



Observations ($d_L=1$):

- ① $A^T A$ depends only on input data,
 $A^+ B$ on both input & output
- ② $A^T A \in \mathbb{R}^{\dim V \times \dim V}$ is rank-deficient whenever $|S| < \dim V \rightarrow$ pseudometric
(LLMs: $|S| < \dim M$)
- ③ even when $|S| \gg \dim V$, $A^T A$ is not an arbitrary symmetric PD matrix,
 while $A^+ B$ yields all vectors $\in \mathbb{R}^{\dim V}$
(why?)
Which matrices can be obtained?
 (try for $d_0=1$: $v(x) = (1, x, x^2, \dots, x^D)$)

$$\arg\min_{f \in \mathcal{M}} L(f) = \arg\min_{f \in \mathcal{M}} \|C_f - A^+ B\|_{A^T A}^2$$

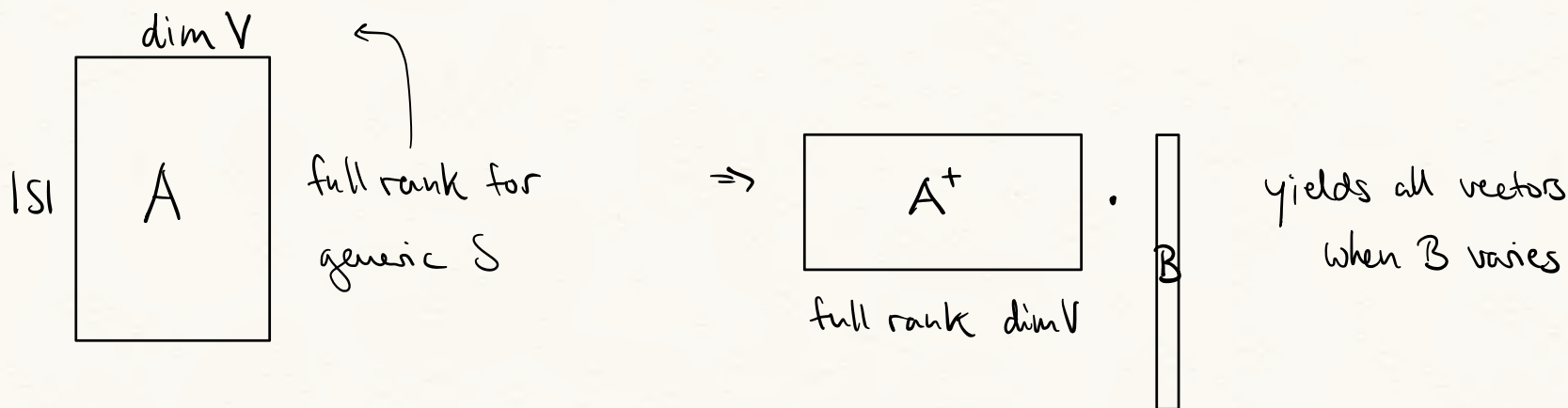


Observations ($d_L=1$):

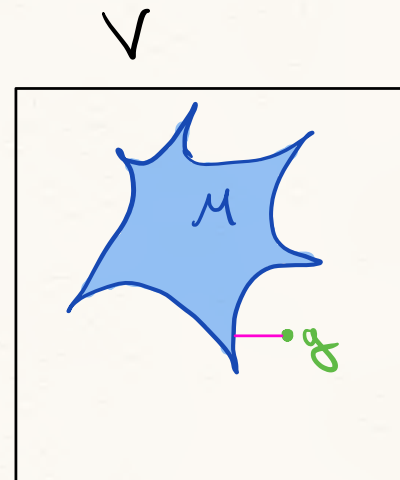
① $A^T A$ depends only on input data,
 $A^+ B$ on both input & output

② $A^T A \in \mathbb{R}^{\dim V \times \dim V}$ is rank-deficient whenever $|S| < \dim V \rightarrow$ pseudometric
 (LLMs: $|S| < \dim M$)

③ even when $|S| \gg \dim V$, $A^T A$ is not an arbitrary symmetric PD matrix,
 while $A^+ B$ yields all vectors $\in \mathbb{R}^{\dim V}$



$$\arg\min_{f \in \mathcal{M}} L(f) = \arg\min_{f \in \mathcal{M}} \|C_f - A^+ B\|_{A^T A}^2$$



Observations ($d_L=1$):

① $A^T A$ depends only on input data,
 $A^+ B$ on both input & output

② $A^T A \in \mathbb{R}^{\dim V \times \dim V}$ is rank-deficient whenever $|S| < \dim V \rightarrow$ pseudometric (LLMs: $|S| < \dim M$)

③ even when $|S| \gg \dim V$, $A^T A$ is not an arbitrary symmetric PD matrix,
 while $A^+ B$ yields all vectors $\in \mathbb{R}^{\dim V}$

$$A^T A = \begin{matrix} & i \rightarrow \\ \begin{array}{|c|c|c|} \hline | & & | \\ \hline v(a_1) & \dots & v(a_{|S|}) \\ \hline | & & | \\ \hline \end{array} & \downarrow i \\ \begin{array}{|c|} \hline v(a_1) \\ \hline \vdots \\ \hline v(a_{|S|}) \\ \hline \end{array} \end{matrix}$$

has (i,j) entry $\sum_{(a,b) \in S} \underbrace{v_i(a) v_j(a)}_{\text{monomial of degree } \leq 2D}$
 that can be factored in several ways

Ex.: $d_0 = 1$

$$\Rightarrow v(x) = (1, x, x^2, \dots, x^D)$$

$$\Rightarrow A = \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^D \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_{|S|} & a_{|S|}^2 & \dots & a_{|S|}^D \end{bmatrix} \quad \text{Vandermonde matrix}$$

$$\Rightarrow A^T A = \begin{bmatrix} |S| & \sum a_k & \sum a_k^2 & \dots & \sum a_k^D \\ \sum a_k & \sum a_k^2 & \sum a_k^3 & \dots & \sum a_k^{D+1} \\ \sum a_k^2 & \sum a_k^3 & \sum a_k^4 & \dots & \sum a_k^{D+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum a_k^D & \sum a_k^{D+1} & \sum a_k^{D+2} & \dots & \sum a_k^{2D} \end{bmatrix} \quad \text{Hankel matrix}$$

Ex.: $d_0 = 1$

$$\Rightarrow v(x) = (1, x, x^2, \dots, x^D)$$

$$\Rightarrow A = \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^D \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_{|S|} & a_{|S|}^2 & \dots & a_{|S|}^D \end{bmatrix} \quad \text{Vandermonde matrix}$$

$$\Rightarrow A^T A = \begin{bmatrix} |S| & \sum a_k & \sum a_k^2 & \dots & \sum a_k^D \\ \sum a_k & \sum a_k^2 & \sum a_k^3 & \dots & \sum a_k^{D+1} \\ \sum a_k^2 & \sum a_k^3 & \sum a_k^4 & \dots & \sum a_k^{D+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum a_k^D & \sum a_k^{D+1} & \sum a_k^{D+2} & \dots & \sum a_k^{2D} \end{bmatrix} \quad \text{Hankel matrix}$$

Ex.: $d_0 = 2, D = 2$

$$\Rightarrow v(x, y) = (1, x, y, x^2, xy, y^2)$$

$$\Rightarrow A^T A = \sum_{\substack{(a,b) \in S \\ a=(x,y)}} \begin{bmatrix} 1 & x & y & x^2 & xy & y^2 \\ 1 & x & y & x^2 & xy & y^2 \\ x & x^2 & xy & x^3 & x^2y & xy^2 \\ y & xy & y^2 & x^2y & xy^2 & y^3 \\ x^2 & x^3 & x^2y & x^4 & x^3y & x^2y^2 \\ xy & x^2y & xy^2 & x^3y & x^2y^2 & xy^3 \\ y^2 & xy^2 & y^3 & x^2y^2 & xy^3 & y^4 \end{bmatrix} \begin{matrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{matrix}$$

Network training = 'distance' minimization

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

$\nwarrow$ neuromanifold

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

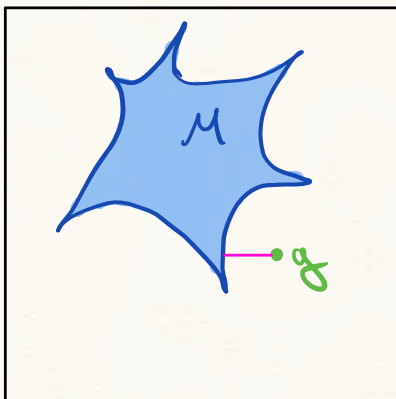
$\nwarrow$ mean squared error

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

$\nwarrow$ [dist(f,g) = 0 possible for $f \neq g$]

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

V



$d_L > 1$

$$f = (f_1, \dots, f_{d_L}), \quad C_f := \begin{bmatrix} | & & | \\ c_{f_1} & \dots & c_{f_{d_L}} \\ | & & | \end{bmatrix}$$

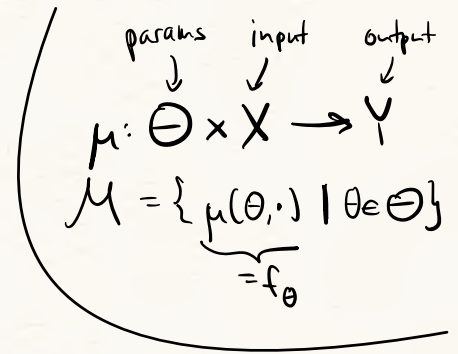
$$\Rightarrow f(x) = v_D(x) \cdot C_f$$

$$\Rightarrow \mathcal{L}(f) = \|A C_f - B\|_{\text{Frob}}^2 = \|C_f - \underbrace{A^+ B}_{\text{ATA}}\|_{\text{ATA}}^2 + \text{const.}$$

$\|C\|_Q^2 := \text{tr}(C^T Q C)$

Loss Landscape

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



Loss Landscape

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

$$\begin{array}{ccc} \text{params} & \text{input} & \text{output} \\ \downarrow & \downarrow & \downarrow \\ \mu: \Theta \times X & \rightarrow & Y \\ \mathcal{M} = \{ \underbrace{\mu(\theta, \cdot)}_{=f_\theta} \mid \theta \in \Theta \} \end{array}$$

can be studied in a decoupled way:

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



loss landscape in function space:

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

Loss Landscape

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

$$\begin{array}{ccc} \text{params} & \text{input} & \text{output} \\ \downarrow & \downarrow & \downarrow \\ \mu: \Theta \times X & \rightarrow & Y \\ \mathcal{M} = \{ \underbrace{\mu(\theta, \cdot)}_{=f_\theta} \mid \theta \in \Theta \} \end{array}$$

can be studied in a decoupled way:

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



loss landscape in function space:

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

How?

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

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

Geometry of Neuron manifolds

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

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

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

A **semialgebraic** set!

describable by
polynomial equations
& inequalities

Geometry of Neuron manifolds

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

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

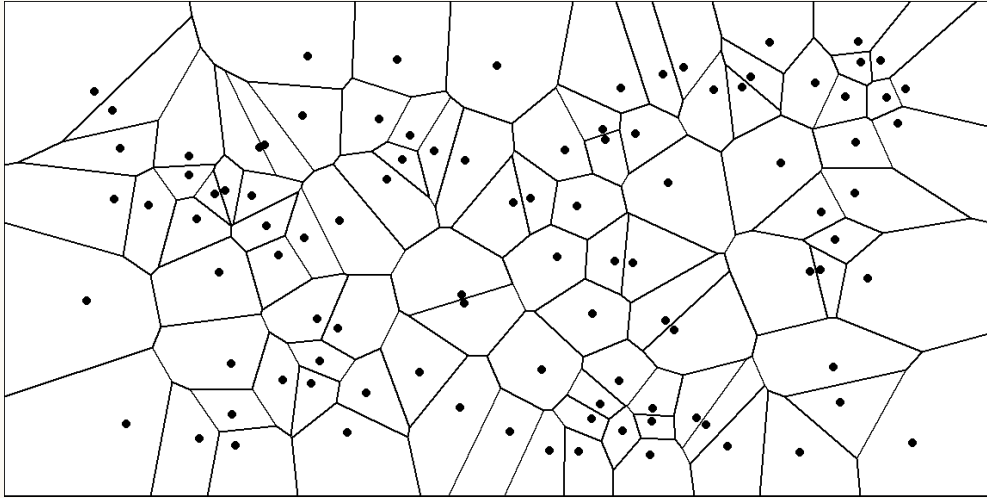
What kind of object is $\mathcal{M}$?

A **semialgebraic** set!

↑
describable by
polynomial equations
& inequalities

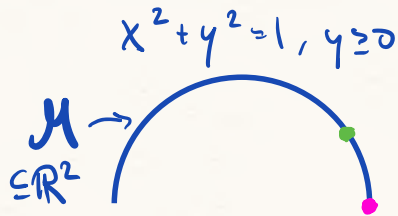
Euclidean distance
minimization can be
implicitly biased to
singularities & boundaries of $\mathcal{M}$

Voronoi cells



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

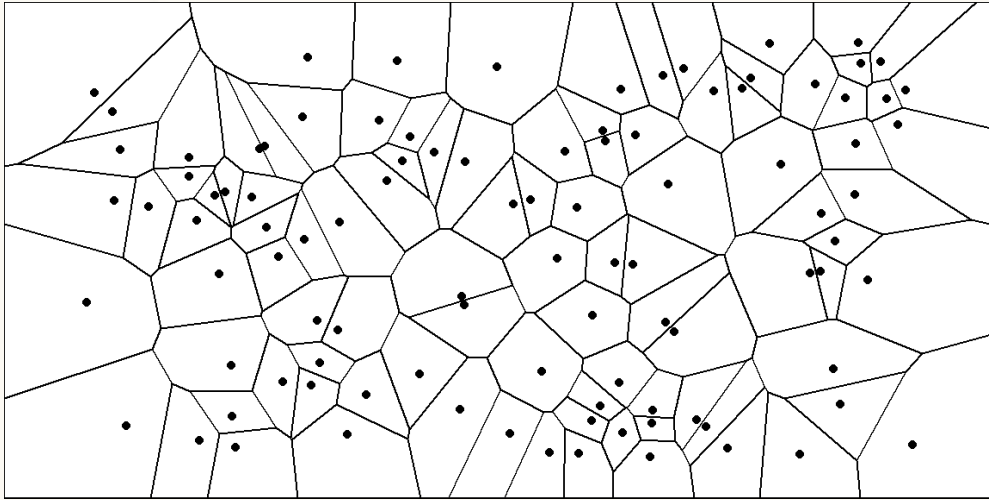
$$\text{Vor}_S(p) := \{u \in \mathbb{R}^n \mid \forall q \in S, q \neq p: \|p - u\|_2 < \|q - u\|_2\}$$



What is the Voronoi cell at •?

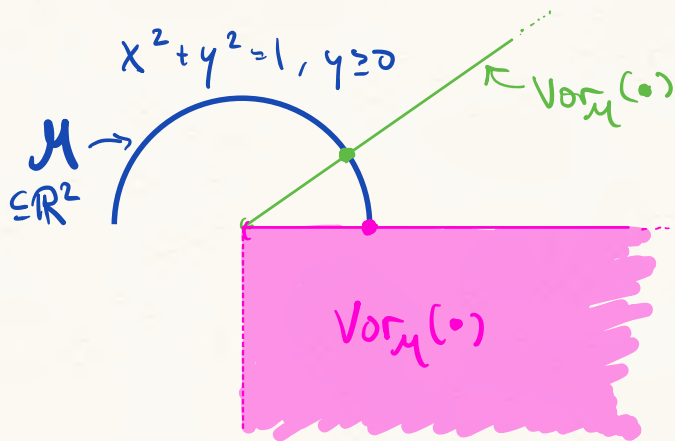
What is the Voronoi cell at •?

Voronoi cells



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

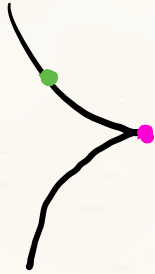
$$\text{Vor}_S(p) := \{u \in \mathbb{R}^n \mid \forall q \in S, q \neq p: \|p - u\|_2 < \|q - u\|_2\}$$



The 2 relative boundary points are the only points on M with full-dimensional Voronoi cells!
 $\Rightarrow$ **implicit bias** towards ∂M

points in ∂M are global minima with positive probability on data u

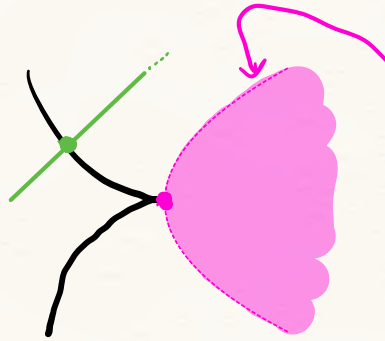
singularities



What are the Voronoi cells at $\bullet$ and $\bullet$?

singularities

$$y^2 + x^3 = 0$$
$$t \mapsto (-t^2, t^3)$$



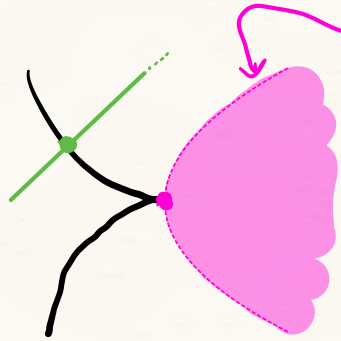
Challenge: Compute this curve!

→ implicit bias towards $\text{Sing}(M)$

What are the Voronoi cells at $\bullet$ and $\bullet$?

singularities

$$y^2 + x^3 = 0$$
$$t \mapsto (-t^2, t^3)$$



Challenge: Compute this curve!

→ **implicit bias** towards $\text{Sing}(M)$

What are the Voronoi cells at $\bullet$ and $\bullet$?

Tradeoff



learning close to singularity
→ slow & numerical instability
[Amari et al]

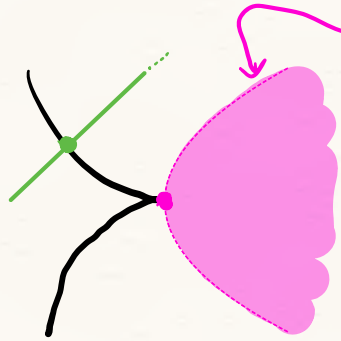


Singular solution generalizes better:

- ① stable global minimum when perturbing data
- ② **Conjecture:** singularities of neuromanifolds are sparse subnetworks
[we've proven this for MPs & CNNs]

singularities

$$y^2 + x^3 = 0$$
$$t \mapsto (-t^2, t^3)$$



Challenge: Compute this curve!

→ **implicit bias** towards $\text{Sing}(M)$

What are the Voronoi cells at $\bullet$ and $\bullet$?

Tradeoff



learning close to singularity
→ slow & numerical instability
[Amaral 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 MIPs & CNMs]

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



MLP

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



CNN

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

In both cases, they are sparse subnetworks :)

What about smooth interior points?

$M \subseteq \mathbb{R}^n$ algebraic variety (i.e. described by polynomial equations)

Q symmetric PD $n \times n$ matrix

Fact: For almost all $u \in \mathbb{R}^n$, the number of complex critical points of

$$\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$$

is the same, called the **Euclidean Distance Degree**: $\text{EDD}_Q(M)$.

What is $\text{EDD}_{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}(\mathcal{O})$?

What about smooth interior points?

$M \subseteq \mathbb{R}^n$ algebraic variety (i.e. described by polynomial equations)

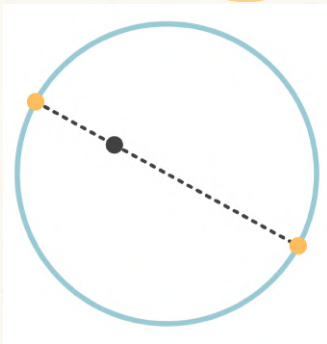
Q symmetric PD $n \times n$ matrix

Fact: For almost all $u \in \mathbb{R}^n$, the number of complex critical points of

$$\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$$

is the same, called the **Euclidean Distance Degree**: $\text{EDD}_Q(M)$.

What is $\text{EDD}_{\begin{bmatrix} 0 & 1 \end{bmatrix}}(\bigcirc)$?



What is $\text{EDD}_{\begin{bmatrix} 4 & 1 \end{bmatrix}}(\bigcirc)$?

What about smooth interior points?

$M \subseteq \mathbb{R}^n$ algebraic variety (i.e. described by polynomial equations)

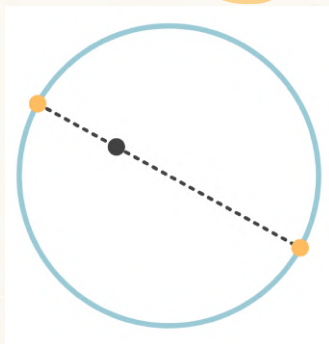
Q symmetric PD $n \times n$ matrix

Fact: For almost all $u \in \mathbb{R}^n$, the number of complex critical points of

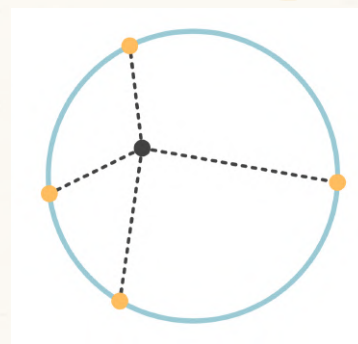
$$\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$$

is the same, called the **Euclidean Distance Degree**: $\text{EDD}_Q(M)$.

What is $\text{EDD}_{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}(\bigcirc)$?



What is $\text{EDD}_{\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}}(\bigcirc)$?



What about smooth interior points?

$M \subseteq \mathbb{R}^n$ algebraic variety (i.e. described by polynomial equations)

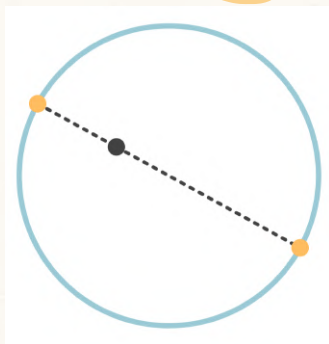
Q symmetric PD $n \times n$ matrix

Fact: For almost all $u \in \mathbb{R}^n$, the number of complex critical points of

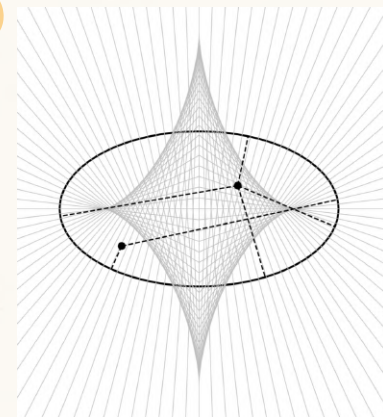
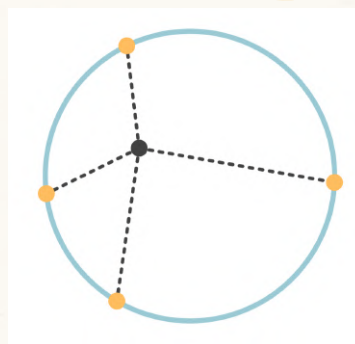
$$\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$$

is the same, called the **Euclidean Distance Degree**: $\text{EDD}_Q(M)$.

What is $\text{EDD}_{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}(\bigcirc)$?



What is $\text{EDD}_{\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}}(\bigcirc)$?

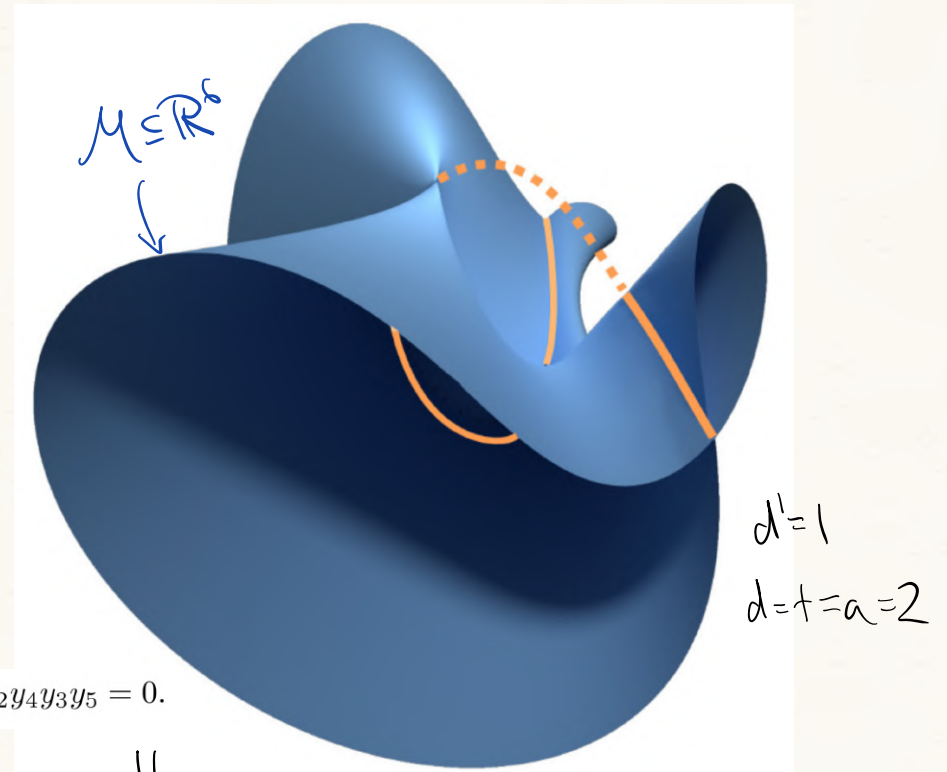


Lightning Self-Attention (single head, single layer)

$$\begin{aligned} \mathbb{R}^{d \times t} &\longrightarrow \mathbb{R}^{d' \times t} \\ X &\longmapsto V X X^T K^T Q X \end{aligned}$$

learnable parameters
 $V \in \mathbb{R}^{d' \times d}$, $K, Q \in \mathbb{R}^{n \times d}$

$$y_1^2 y_6^2 + y_4^2 y_3^2 + y_1 y_3 y_5^2 + y_2^2 y_4 y_6 - 2 y_1 y_4 y_3 y_6 - y_2 y_1 y_6 y_5 - y_2 y_4 y_3 y_5 = 0.$$



$\Downarrow$
 For almost all PD matrices Q ,
 $\text{EDD}_Q(M) = 14$.

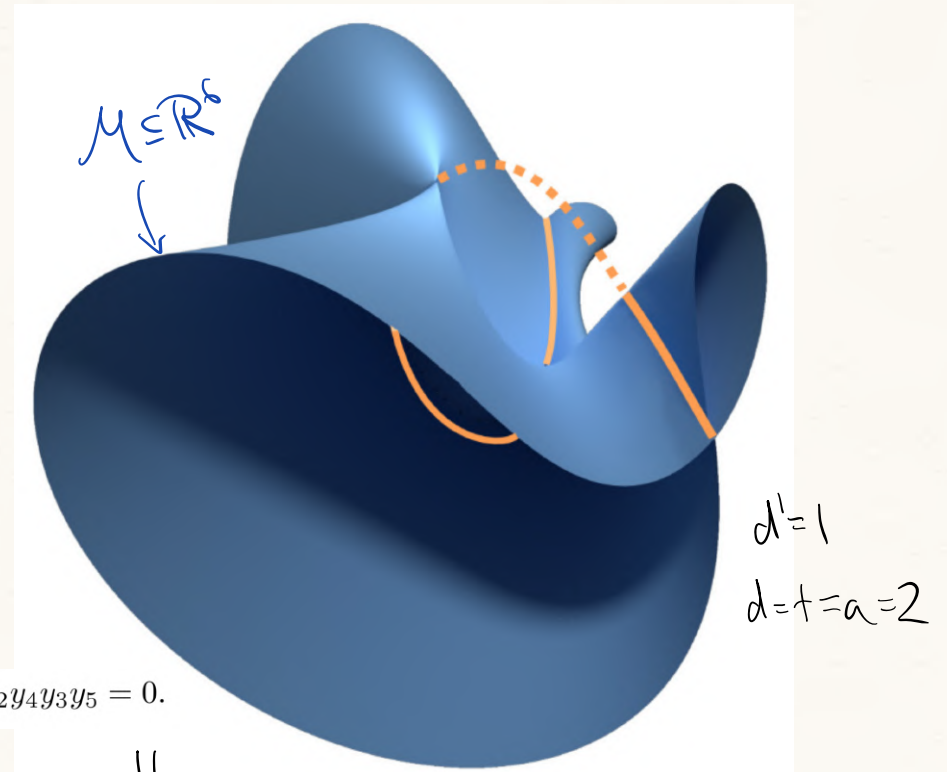
What happens if Q becomes degenerate?
 (ie., Q is symmetric positive semidefinite)

Lightning Self-Attention (single head, single layer)

$$\begin{aligned} \mathbb{R}^{d \times t} &\longrightarrow \mathbb{R}^{d' \times t} \\ X &\longmapsto V X X^T K^T Q X \end{aligned}$$

learnable parameters
 $V \in \mathbb{R}^{d' \times d}$, $K, Q \in \mathbb{R}^{n \times d}$

$$y_1^2 y_6^2 + y_4^2 y_3^2 + y_1 y_3 y_5^2 + y_2^2 y_4 y_6 - 2 y_1 y_4 y_3 y_6 - y_2 y_1 y_6 y_5 - y_2 y_4 y_3 y_5 = 0.$$



For almost all PD matrices Q ,
 $\text{EDD}_Q(M) = 14$.

What happens if Q becomes degenerate?
 (ie., Q is symmetric positive semidefinite)

k	complex critical point set
0	14 points
1	14 points
2	4 points + a curve
3	a surface
4	a 3-dimensional subvariety
5	a 4-dimensional subvariety

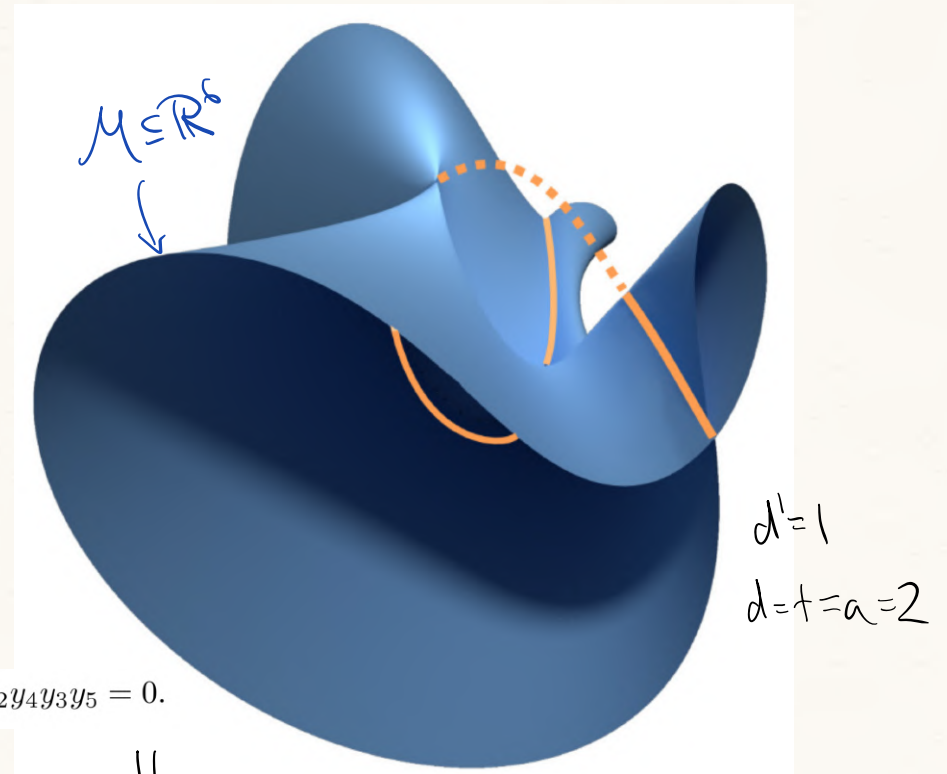
$k := \dim \ker Q$

Lightning Self-Attention (single head, single layer)

$$\begin{aligned} \mathbb{R}^{d \times t} &\longrightarrow \mathbb{R}^{d' \times t} \\ X &\longmapsto V X X^T K^T Q X \end{aligned}$$

learnable parameters
 $V \in \mathbb{R}^{d' \times d}$, $K, Q \in \mathbb{R}^{n \times d}$

$$y_1^2 y_6^2 + y_4^2 y_3^2 + y_1 y_3 y_5^2 + y_2^2 y_4 y_6 - 2 y_1 y_4 y_3 y_6 - y_2 y_1 y_6 y_5 - y_2 y_4 y_3 y_5 = 0.$$



For almost all PD matrices Q ,
 $\text{EDD}_Q(M) = 14$.

What happens if Q becomes degenerate?
 (ie., Q is symmetric positive semidefinite)

k	complex critical point set
0	14 points
1	14 points
2	4 points + a curve
3	a surface
4	a 3-dimensional subvariety
5	a 4-dimensional subvariety

$k := \dim \ker Q$

$M \cap (\ker(Q) + u)$
 $\hookrightarrow$ zero loss solutions!

In general

$M \subseteq \mathbb{R}^n$ algebraic variety, $d := \dim M$.

Q symmetric positive semi-definite $n \times n$ matrix
 $K := \ker Q$

$$\pi: \mathbb{R}^n \rightarrow K^\perp$$

turns Q into nondegenerate quadric

In general

$M \subseteq \mathbb{R}^n$ algebraic variety, $d := \dim M$.

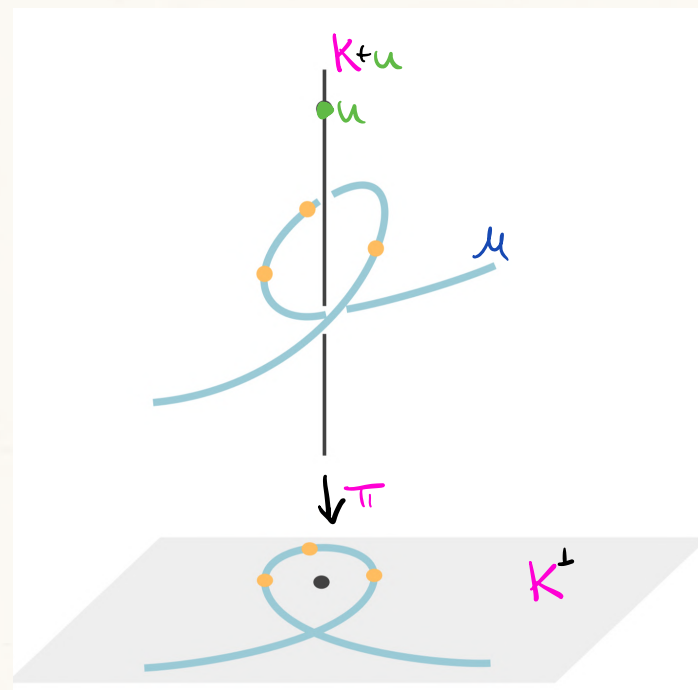
Q symmetric positive semi-definite $n \times n$ matrix
 $K := \ker Q$

$\pi: \mathbb{R}^n \rightarrow K^\perp$
 turns Q into nondegenerate quadric

Case I: Let $k < n - d$.

For almost all Q with $k = \dim K$ and almost all $u \in \mathbb{R}^n$,

$$\begin{array}{l} \text{EDD}_Q(M) \\ \parallel \\ \text{EDD}_{\pi(Q)}(\pi(M)) \end{array} \left\{ \begin{array}{l} \text{critical points of } \min_{x \in M} \|x - u\|_Q^2 \\ \updownarrow 1:1 \\ \text{critical points of } \min_{x \in \pi(M)} \|x - \pi(u)\|_{\pi(Q)}^2 \end{array} \right.$$



In general

$M \subseteq \mathbb{R}^n$ algebraic variety, $d := \dim M$.

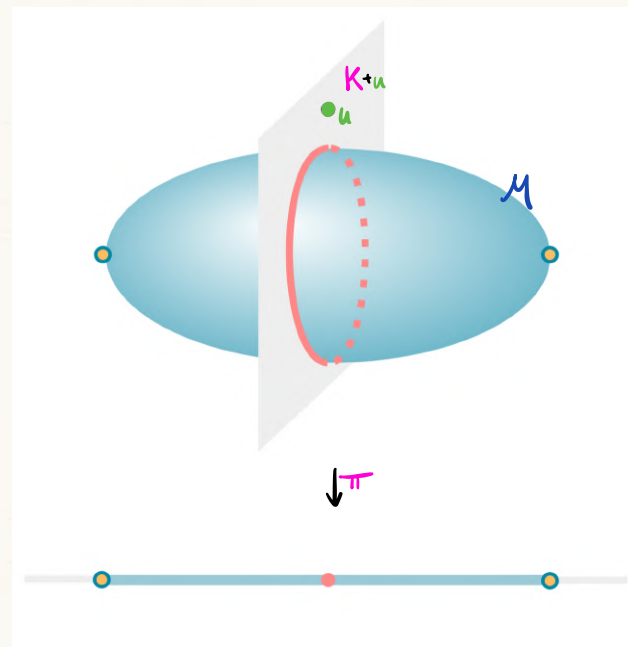
Q symmetric positive semi-definite $n \times n$ matrix
 $K := \ker Q$

$\pi: \mathbb{R}^n \rightarrow K^\perp$
 turns Q into nondegenerate quadric

Case 2: Let $k \geq n - d$.

For almost all Q with $k = \dim K$ and almost all $u \in \mathbb{R}^n$, we have
 2 types of critical points of $\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$:

① $(K+u) \cap M$: zero loss solutions



In general

$M \subseteq \mathbb{R}^n$ algebraic variety, $d := \dim M$.

Q symmetric positive semi-definite $n \times n$ matrix
 $K := \ker Q$

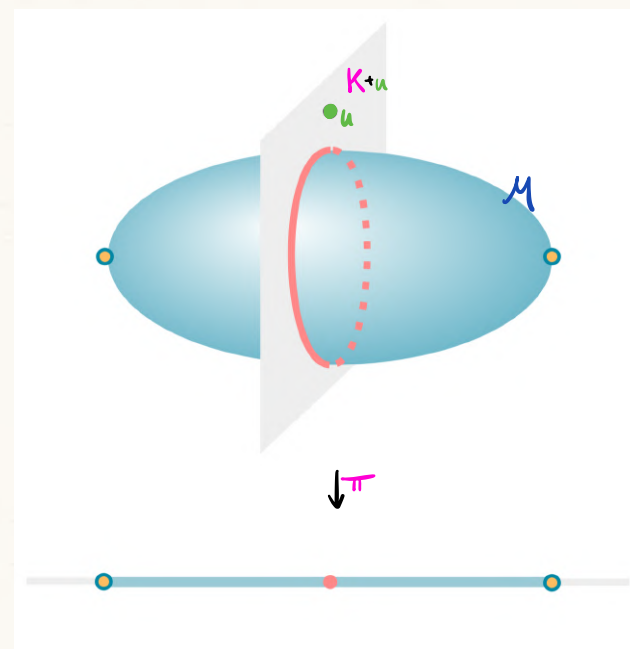
$\pi: \mathbb{R}^n \rightarrow K^\perp$
 turns Q into nondegenerate quadric

Case 2: Let $k \geq n - d$.

For almost all Q with $k = \dim K$ and almost all $u \in \mathbb{R}^n$, we have
 2 types of critical points of $\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$:

Ⓐ $(K+u) \cap M$: zero loss solutions

Ⓑ finitely many on the ramification locus $\text{Ram}(\pi|_M)$
 $:= \{\text{critical points of } \pi|_M\}$



In general

$M \subseteq \mathbb{R}^n$ algebraic variety, $d := \dim M$.

Q symmetric positive semi-definite $n \times n$ matrix
 $K := \ker Q$

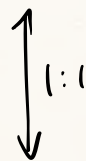
$\pi: \mathbb{R}^n \rightarrow K^\perp$
 turns Q into nondegenerate quadric

Case 2: Let $k \geq n - d$.

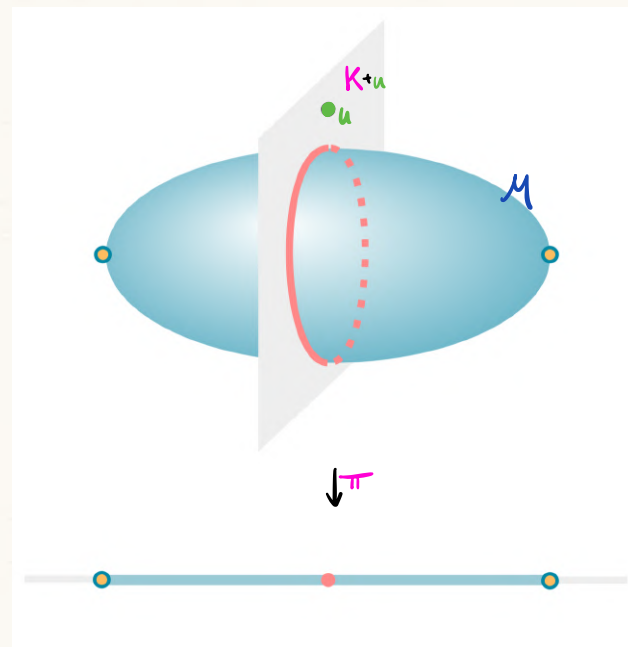
For almost all Q with $k = \dim K$ and almost all $u \in \mathbb{R}^n$, we have
 2 types of critical points of $\min_{x \in M \setminus \text{Sing}(M)} \|x - u\|_Q^2$:

Ⓐ $(K+u) \cap M$: zero loss solutions

Ⓑ finitely many on the ramification locus $\text{Ram}(\pi|_X)$
 $:= \{\text{critical points of } \pi|_X\}$



$\text{EDD}_{\pi(Q)}(\text{Br}) \leftarrow \text{critical points of } \min_{x \in \text{Br}(\pi|_X)} \|x - \pi(u)\|_{\pi(Q)}^2$
 $\text{Br}(\pi|_X) \leftarrow \text{Branch locus } \pi(\text{Ram})$



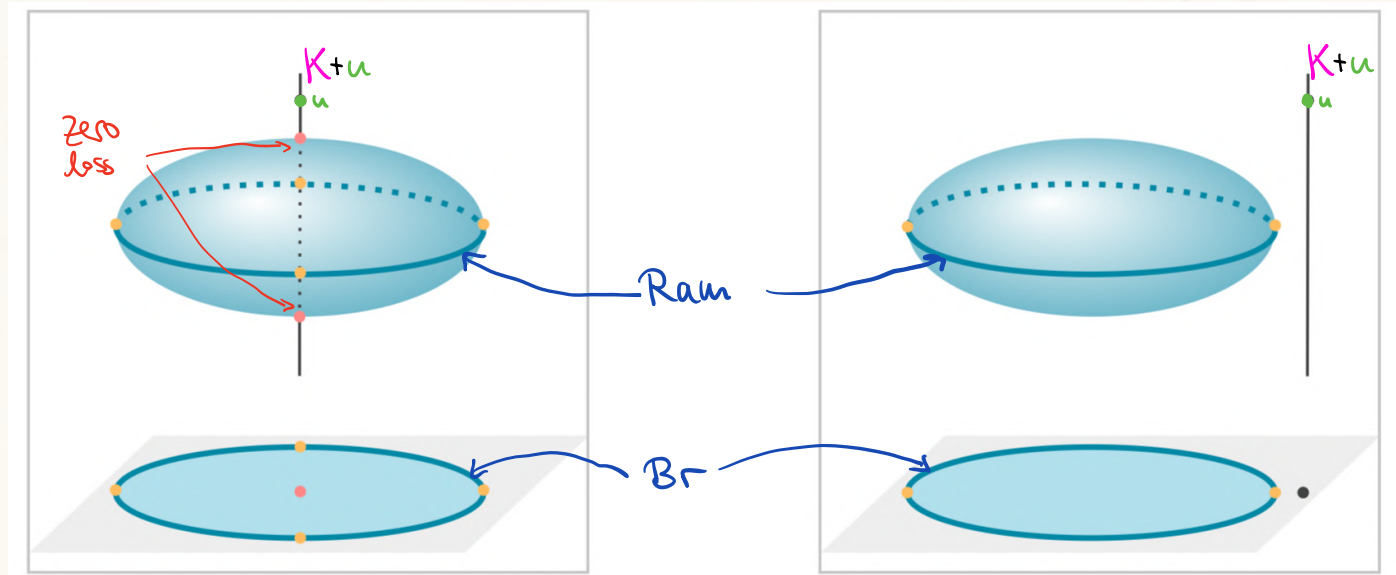
In general

$M \subseteq \mathbb{R}^n$ algebraic variety, $d := \dim M$.

Q symmetric positive semi-definite $n \times n$ matrix
 $K := \ker Q$

$\pi: \mathbb{R}^n \rightarrow K^\perp$
 turns Q into nondegenerate quadric

Case 2: let $k \geq n-d$.



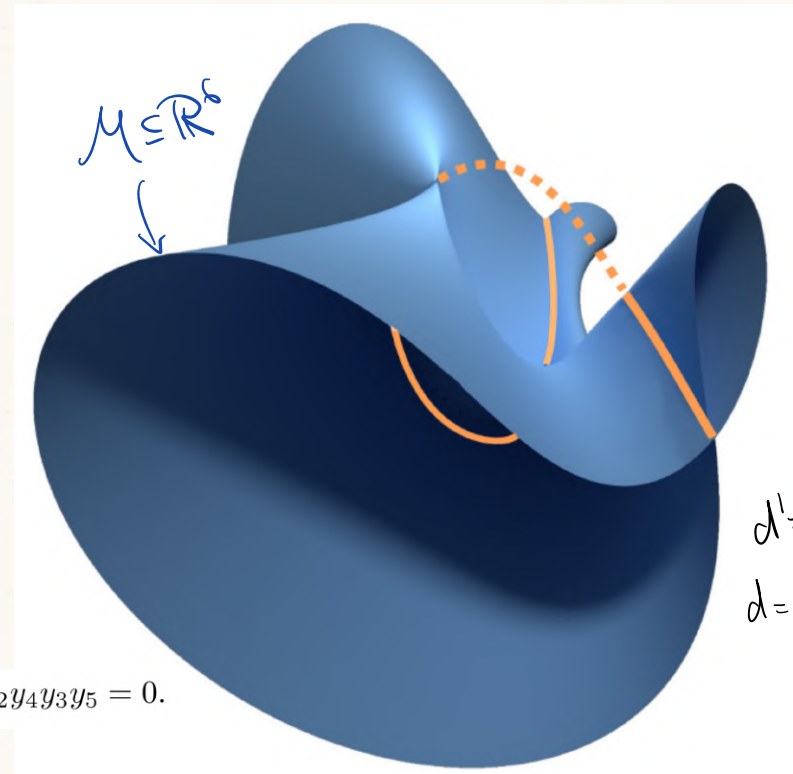
Induced bias towards Ram!

depends only on K (not on Q) & not on u

Lightning Self-Attention (single head, single layer)

$$\begin{aligned} \mathbb{R}^{d \times t} &\longrightarrow \mathbb{R}^{d' \times t} \\ X &\longmapsto V X X^T K^T Q X \end{aligned}$$

learnable parameters
 $V \in \mathbb{R}^{d' \times d}$, $K, Q \in \mathbb{R}^{n \times d}$



$$y_1^2 y_6^2 + y_4^2 y_3^2 + y_1 y_3 y_5^2 + y_2^2 y_4 y_6 - 2 y_1 y_4 y_3 y_6 - y_2 y_1 y_6 y_5 - y_2 y_4 y_3 y_5 = 0.$$

Q = from MSE loss with general dataset S

$ S $	$k = \dim K$	complex critical point set
≥ 3	0	14 points
2	2	a curve and two lines
1	4	a 3-dimensional subvariety

Q = general symmetric positive semidefinite

k	complex critical point set
0	14 points
1	14 points
2	4 points + a curve
3	a surface
4	a 3-dimensional subvariety
5	a 4-dimensional subvariety

$K := \dim \ker Q$

$M \cap (\ker(Q) + u)$
 $\hookrightarrow$ zero loss solutions!

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}

Machine Learning	Algebraic Geometry
sample complexity and expressivity	dimension, degree, and covering number
subnetworks and implicit bias	singularities
identifiability and invariance	fibers of the parameterization
optimization and gradient descent	critical point theory, discriminants, and dynamical invariants

Identifiability

$$\begin{array}{ccc} \Theta & \xrightarrow{\mu} & \mathcal{M} \\ \theta & \mapsto & f_\theta \end{array}$$

Given (generic) $f \in \mathcal{M}$,
what is $\hat{\mu}(f)$?

Identifiability

$$\begin{array}{ccc} \Theta & \xrightarrow{\mu} & \mathcal{M} \\ \theta & \longmapsto & f_\theta \end{array}$$

Given (generic) $f \in \mathcal{M}$,
what is $\bar{\mu}^{-1}(f)$?

For monomial MLP with $\sigma(x) = x^r, r \gg 0$:

$$\mu: \mathbb{R}^{d_L \times d_{L-1}} \times \dots \times \mathbb{R}^{d_1 \times d_0} \longrightarrow \mathcal{M}$$

$$(W_L, \dots, W_1) \longmapsto W_L \circ \sigma \circ \dots \circ \sigma \circ W_2 \circ \sigma \circ W_1$$

What is the generic fiber?

Identifiability

$$\begin{array}{ccc} \Theta & \xrightarrow{\mu} & \mathcal{M} \\ \theta & \longmapsto & f_\theta \end{array}$$

Given (generic) $f \in \mathcal{M}$,
what is $\hat{\mu}(f)$?

For monomial MLP with $\sigma(x) = x^r, r \gg 0$:

$$\mu: \mathbb{R}^{d_L \times d_{L-1}} \times \dots \times \mathbb{R}^{d_1 \times d_0} \longrightarrow \mathcal{M}$$

$$(W_L, \dots, W_1) \longmapsto W_L \circ \sigma \circ \dots \circ \sigma \circ W_2 \circ \sigma \circ W_1$$

Observation: $D_i \in GL(d_i)$ diagonal

$P_i \in GL(d_i)$ permutation matrix

$$\begin{aligned} \Rightarrow \mu(W_L D_{L-1}^{-r} P_{L-1}^T, \dots, P_2 D_2 W_2 D_1^{-r} P_1^T, P_1 D_1 W_1) \\ = \mu(W_L, \dots, W_2, W_1) \end{aligned}$$

On the Expressive Power of
Deep Polynomial Neural Networks ²⁰¹⁹

Joe Kileel*
Princeton University

Matthew Trager*
New York University

Joan Bruna
New York University

Identifiability

$$\begin{array}{ccc} \Theta & \xrightarrow{\mu} & \mathcal{M} \\ \theta & \mapsto & f_\theta \end{array}$$

Given (generic) $f \in \mathcal{M}$,
what is $\bar{\mu}(f)$?

For monomial MLP with $\sigma(x) = x^r$, $r \gg 0$:

Activation degree thresholds and expressiveness of
polynomial neural networks 2024

Bella Finkel*, Jose Israel Rodriguez†, Chenxi Wu, Thomas Yahl

Proven that those parameter symmetries are
the generic fiber
(implicitly described all fibers!)

follow-ups (2025):

THE ALEXANDER-HIRSCHOWITZ THEOREM FOR NEUROVARIETIES

ALEX MASSARENTI AND MASSIMILIANO MELLA

Identifiability of Deep Polynomial Neural Networks

Konstantin Usevich*, Ricardo Borsoi, Clara Dérand, Marianne Clausel†
Université de Lorraine, CNRS, CRAN

Identifiability

$$\begin{array}{ccc} \Theta & \xrightarrow{\mu} & \mathcal{M} \\ \theta & \mapsto & f_\theta \end{array}$$

Given (generic) $f \in \mathcal{M}$,
what is $\bar{\mu}(f)$?

For monomial MLP with $\sigma(x) = x^r$, $r \gg 0$:

Activation degree thresholds and expressiveness of
polynomial neural networks 2024

Bella Finkel*, Jose Israel Rodriguez†, Chenxi Wu, Thomas Yahl

Proposition 16. Let $\mathbb{K}$ be a subfield of $\mathbb{C}$. Given integers d, k , there exists an integer $\tilde{r} = \tilde{r}(k)$ with the following property. If $r > \tilde{r}(k)$ and $p_1, \dots, p_k \in \mathbb{K}[x_1, \dots, x_d]$ are pairwise non-proportional, then $p_1^r, \dots, p_k^r$ are linearly independent (over $\mathbb{K}$). Moreover, $\tilde{r}(k) = 6(k-1)^2 - 6(k-1) + 1$ has the desired property.

Identifiability

$$\begin{array}{ccc} \Theta & \xrightarrow{\mu} & \mathcal{M} \\ \theta & \longmapsto & f_\theta \end{array}$$

Given (generic) $f \in \mathcal{M}$,
what is $\bar{\mu}'(f)$?

For polynomial MLP with $\sigma(x) \in \mathbb{R}[x]_{\leq r}$ generic, $r \geq 0$:

Conjecture: Generic fiber $\bar{\mu}'(f)$ consists only of permutations.

Conjecture: Let $d, k \in \mathbb{Z}_{>0}$.

There is $\tilde{r} \in \mathbb{Z}_{>0}$ such that all $r \geq \tilde{r}$ satisfy:

There is $U \subseteq \mathbb{R}[x]_{\leq r}$ Zariski open such that:

For all $\sigma \in U$ and all $p_1, \dots, p_k \in \mathbb{R}[x_1, \dots, x_d]$ non-constant & pairwise distinct:

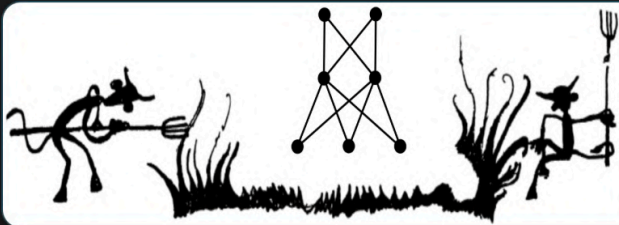
$\sigma(p_1), \dots, \sigma(p_k)$ are linearly independent.



Neuroalgebraic Geometry

Auto

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

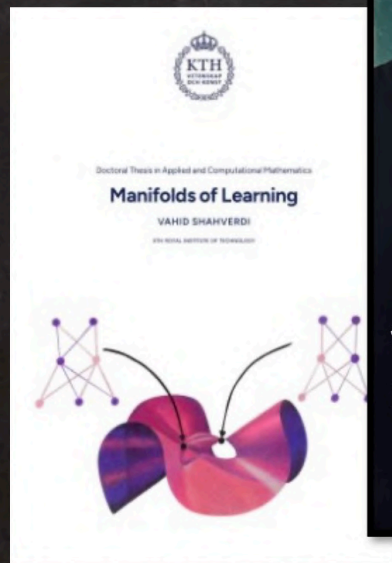
Geometry and Optimization of Shallow Polynomial Networks

Arjevani, Bruna, Kileel, Polak, Trager

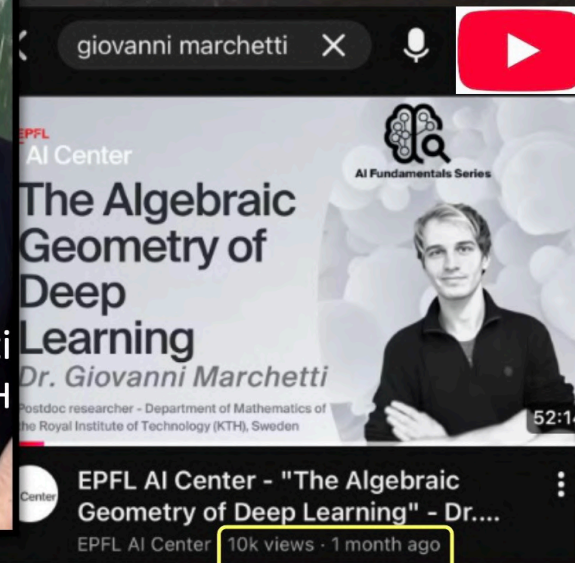
SIAM Journal on Applied Algebra and Geometry · 2026

[Read paper](#)

special thanks to



Giovanni Marchetti
KTH



Matthew Trager
AWS AI Labs, NY

Stefano Mereta
CUNEF Madrid

Joan Bruna
Courant Institute, NYU

Nathan Henry
UC Berkeley

Guido Montúfar
UCLA & MPI MiS Leipzig

Paul Bröding & Erin Connelly
Univ. Osnabrück

Announcements:

I will give a **course at the summer school** of the Italian Network for Applied and Birational Algebraic Geometry (INABAG) in Siena: apply for funding by April 17

Open **PhD position** in my group: apply by May 20