Metric Algebraic Geometry Tutorial

Paul Breiding & Kathlén Kohn

SIAM AG 25

July 7, 2025

Teaser – Training Neural Networks

A Shallow Neural Network

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

- the activation function $\sigma(X) = X^4$ gets applied entrywise
- $a, b, \ldots, f$ are the learnable parameters

This parametrizes quartic homogeneous polynomials in (x, y):

$$Ax^4 + Bx^3y + Cx^2y^2 + Dxy^3 + Ey^4.$$

The Zariski closure of the set of all parametrized polynomials is a 3-fold in $\mathbb{P}^4$:

$$2C^3 - 9BCD + 27AD^2 + 27B^2E - 72ACE = 0.$$

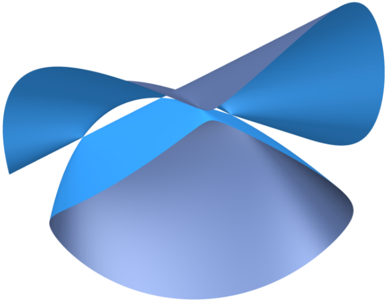


Figure: $C = 1, A + B = D + E$

Neuromanifold & Network Training

$$(a, b, \dots, f) \longmapsto \mu(x, y) = \begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right) \in \mathrm{Sym}_4(\mathbb{R}^2)$$

The image of this map is a proper semi-algebraic set, called the **neuromanifold** $\mathcal{M}$ of the network (although it has singularities!)

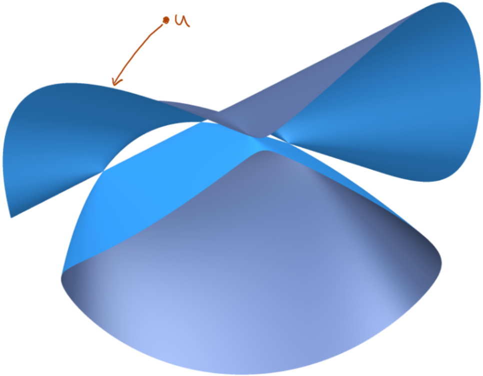
Let's train the network by minimizing the mean squared error loss for given training data $\mathcal{D} = \{(x_1, y_1, z_1), \dots, (x_1, y_1, z_d)\}$:

$$\arg\min_{\mu \in \mathcal{M}} \sum_{i=1}^{d} (z_i - \mu(x_i, y_i))^2$$

Distance Minimization on Neuromanifold

Proposition:

$$\arg\min_{\mu\in\mathcal{M}} \sum_{i=1}^{d} (z_i - \mu(x_i, y_i))^2 = \arg\min_{\mu\in\mathcal{M}} (\mu - u)^\top Q(\mu - u), \quad \text{where}$$



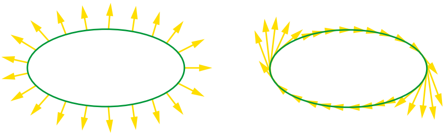
$$Q := V^\top V,\ u := V^+ z,$$

$$V := \begin{bmatrix} x_1^4 & x_1^3 y_1 & x_1^2 y_1^2 & x_1 y_1^3 & y_1^4 \\ x_2^4 & x_2^3 y_2 & x_2^2 y_2^2 & x_2 y_2^3 & y_2^4 \\ & & \vdots & & \\ x_d^4 & x_d^3 y_d & x_d^2 y_d^2 & x_d y_d^3 & y_d^4 \end{bmatrix}$$

Curvature & Volumes of Tubes

Plane Curves & Curvature

- Let $C = \{f(x_1, x_2) = 0\} \subset \mathbb{R}^2$, $\nabla f(x) \neq 0$ on C.
- Unit normal and tangent fields:

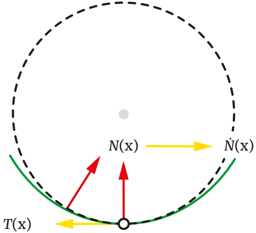


$$N(x) = \frac{\nabla f(x)}{\|\nabla f(x)\|}, \qquad T(x) = (N_2(x),\, -N_1(x)).$$

- Signed curvature

$$c(x) \;=\; \langle T,\; T_1\, \partial_{x_1} N + T_2\, \partial_{x_2} N \rangle = \frac{T^T\, H\, T}{\|\nabla f\|},$$

where H is the Hessian of f.



Regions of high curvature are often critical points of distance minimization!

- Radius of curvature $r(x) = 1/c(x)$,
  center of curvature

$$\Gamma(x) = x - r(x)\, N(x).$$

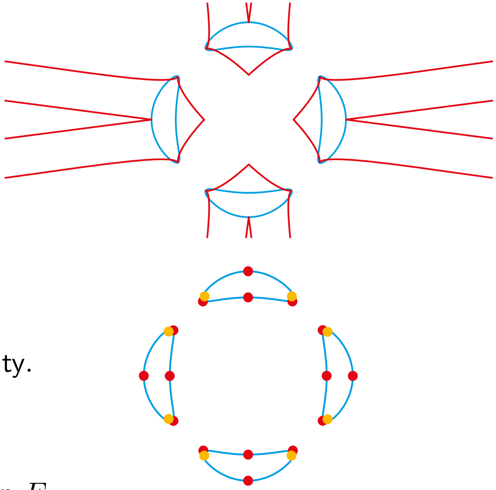
- The *evolute / ED discriminant* E is the Zariski-closure of all centers $\Gamma(x)$.
- Special points on C:

  Inflection point:

$$c(x) = 0 \;\Leftrightarrow\; \Gamma(x) \text{ at infinity.}$$

  Critical curvature:

$$\nabla c(x) \perp T(x) \;\Leftrightarrow\; \text{ cusp on } E.$$



On the ED discriminant, critical points of Euclidean distance collide.

Counting Inflection & Critical Points

- Homogenize $f \to F(x_0, x_1, x_2)$. Let H_0 be its 3×3 Hessian.
- Curvature formula

$$c(x) = \frac{-\det H_0}{(d-1)^2\,(f_1^2 + f_2^2)^{3/2}}\Big|_{x_0=1}.$$

- Inflection points: $f = \det H_0 = 0$.

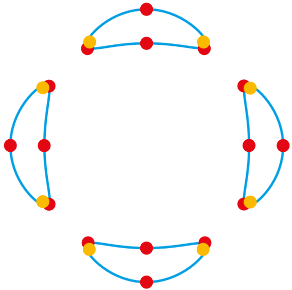
By Bézout: $\#_{\mathbb{C}} = 3d(d-2)$,

By Klein: $\#_{\mathbb{R}} \leq d(d-2)$.

- Critical curvature:

$$\#_{\mathbb{C}} = 2d(3d-5).$$

- Example (Trott curve, $d = 4$): 8 real inflections, 24 real critical points.

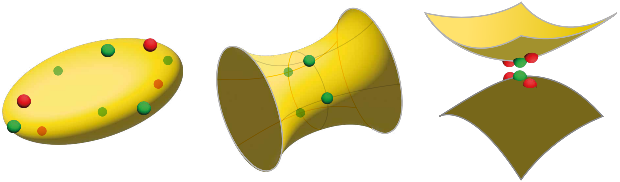


Curvature of Higher-Dimensional Varieties

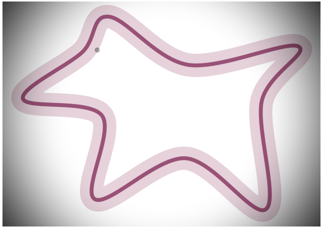
- Let $X \subset \mathbb{R}^n$ be cut out by $f_1, \ldots, f_k$, Jacobian $J = \left(\nabla f_1(x) \cdots \ \nabla f_k(x) \right)$.
- A normal vector $v = J\,w \neq 0$, unit normal $N = v/\|v\|$. Tangent $t \in T_x X$.
- *Curvature in direction* (t, v):

$$c(x,t,v) \;=\; \frac{1}{\|v\|}\, t^T\Big(\sum_{i=1}^{k} w_i\, H_i\Big)\, t.$$

- This quadratic form on $T_x X$ is the *second fundamental form* II_v.
- Its self-adjoint linear map is the *Weingarten map* L_v.
  Eigenvalues $=$ principal curvatures.



Volumes of Tubular Neighborhoods



Tube of radius ε:

$$\mathrm{Tube}(X,\varepsilon) = \{u \in \mathbb{R}^n \mid \min_{x\in X} \|x-u\| < \varepsilon\}.$$

For X a neuromanifold, the volume of the tube measures the expressivity of the neural network!

Let $X \subset \mathbb{R}^n$ be smooth and compact.

- The *reach* of X is the supremum over all $\varepsilon > 0$ such that the exponential map

  $$\varphi_\varepsilon : \mathcal{N}_\varepsilon X = \{(x,v) \mid x \in X, v \perp T_x X, \|v\| < \varepsilon\} \to \mathrm{Tube}(X,\varepsilon),\ (x,v) \mapsto x+v$$

  is a diffeomorphism.
- For $\varepsilon <$ the reach of X: *Weyl's tube formula:*

  $$\mathrm{vol}\big(\mathrm{Tube}(X,\varepsilon)\big) = \sum\nolimits_{0\le 2i\le m} \kappa_{2i}(X)\,\varepsilon^{\,n-m+2i}, \quad m = \dim(X),$$

  where κ_{2i} are integrals of the $2i$-minors of the Weingarten map L_w.

Medial Axis & Offset

Medial Axis

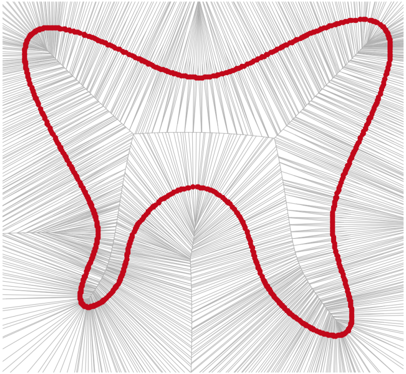
The *medial axis* $\mathrm{Med}(X) \subset \mathbb{R}^n$ is the set of points having *at least two* distinct closest points on X.

If X is semialgebraic then so is $\mathrm{Med}(X)$.

Proposition:

$$\mathrm{dist}\big(X, \mathrm{Med}(X)\big) \;=\; \mathrm{reach}(X).$$

Hence points within distance $< \mathrm{reach}(X)$ from X have a unique nearest point on X.



Bottlenecks, Curvature, and Reach

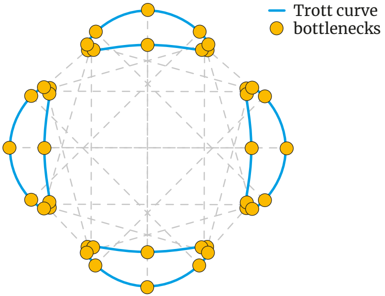
- A *bottleneck* is a pair $\{x, y\} \subset X$, $x \neq y$, for which $x - y$ is normal to both T_xX and T_yX.
- Its *width* is $b(x, y) = \frac{1}{2}\|x - y\|$.

$$B(X) = \min_{\text{bottlenecks}} b(x, y).$$

- The *maximal curvature* of X is

$$C(X) = \max_{x \in X} \max_{i} c_i(x),$$

where $c_i(x)$ are principal curvatures at x.



Theorem: For X smooth,

$$\mathrm{reach}(X) = \min\{B(X), 1/C(X)\}.$$

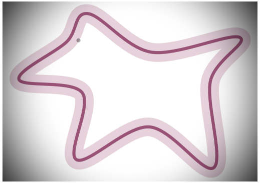
Offset Hypersurfaces & Offset Polynomial

- Let $X \subset \mathbb{R}^n$ be irreducible. Its *ED correspondence* is

$$\mathcal{E}_X = \overline{\{(x,u) \mid x \in X,\ u - x \perp T_x X\}} \subset X \times \mathbb{C}^n.$$

- *Offset correspondence*:

$$\mathcal{OC}_X = \{(x,u,\varepsilon) \in \mathcal{E}_X \times \mathbb{C} \mid \|u-x\|^2 = \varepsilon^2\}.$$



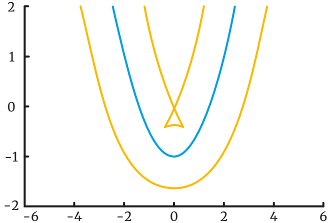
- The closure of its projection to (u,ε) is the *offset hypersurface*

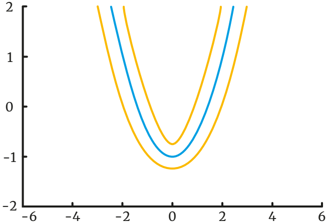
$$\mathrm{Off}_X \subset \mathbb{C}^n \times \mathbb{C}, \quad \mathrm{codim} = 1.$$

- Hence there is a defining *offset polynomial*

$$g_X(u,\varepsilon) = 0.$$

Offset Hypersurface of the Parabola





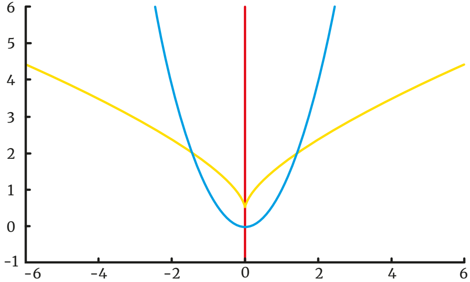
Offset Discriminant & its Decomposition

Define the *offset discriminant* $\delta_X(u) = \mathrm{Disc}_\varepsilon\big(g_X(u,\varepsilon)\big)$ and

$$\Delta_X^{\mathrm{Off}} = V(\delta_X) \subset \mathbb{C}^n.$$

- A point u lies in Δ_X^{Off} iff
  - it has a multiple critical value ($u \in \Sigma_X$, the ED discriminant),
  - or two distinct critical points lie at equal distance (the bisector hypersurface Bis_X).
- Theorem (Horobeț–Weinstein): Write $M_X := \overline{\mathrm{Med}(X)}$. Then

$$\Delta_X^{\mathrm{Off}} = \mathrm{Bis}_X \cup \Sigma_X \;\supseteq\; X \cup M_X \cup \Sigma_X.$$



Computing Normals & Curvature from the Offset Polynomial

- For $u \notin \Delta_X^{\mathrm{Off}}$, let $\varepsilon(u)$ be a local real root of $g_X(u,\varepsilon) = 0$. Suppose that $x \in X$ is the critical point corresponding to (u,ε). By implicit differentiation,

$$\nabla_u \varepsilon(u) = -\left(\tfrac{\partial g_X}{\partial \varepsilon}\right)^{-1} \tfrac{\partial g_X}{\partial u},$$

which is a unit normal vector at x on X.

- Differentiating $\nabla_u \varepsilon(u)$ in direction $t \in T_x X$ gives the second fundamental form evaluated at t. This means:

$$\mathrm{II}_{u-x}(t) \;=\; \lim_{\substack{s \to 0 \\ s > 0}} \; t^\top \left(\frac{\partial^2 \varepsilon}{\partial u^2}(x + s(u-x),\; s\varepsilon) \right) t.$$

- **Conclusion:** from g_X one extracts both the normal field and all principal curvatures of X.

example: parabola

For $X = V(x_2 - x_1^2)$, we find $\frac{d\varepsilon}{du}(u,\varepsilon) = \frac{1}{p}(h_1, h_2)$, where

$$h_1 = -96u_1\varepsilon^4 + \left(192u_1^3 + 64u_1u_2^2 - 16u_1u_2 + 40u_1\right)\varepsilon^2 - 4u_1\left(u_1^2 - u_2\right)\left(24u_1^2 + 16u_2^2 - 16u_2 + 1\right)$$

$$\begin{aligned} h_2 = {} & (-32u_2 - 32)\,\varepsilon^4 + \left(64u_1^2u_2 - 8u_1^2 + 96u_2^2 + 16u_2 - 8\right)\varepsilon^2 \\ & - 2\left(u_1^2 - u_2\right)\left(16u_1^2u_22 - 20u_1^2 - 32u_2^2 + 12u_2 - 1\right) \end{aligned}$$

$$\begin{aligned} p = {} & -96\varepsilon^5 + \left(192u_1^2 + 64u_2^2 + 128u_2 - 32\right)\varepsilon^3 \\ & + \left(-96u_1^4 - 64u_1^2u_2^2 + 16u_1^2u_2 - 64u_2^3 - 40u_1^2 - 16u_2^2 + 16u_2 - 2\right)\varepsilon. \end{aligned}$$

example:

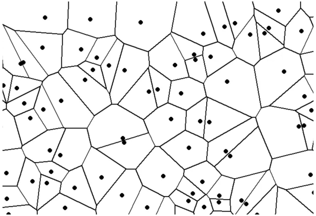
- for $u = (0, \frac{1}{4})$ and $\varepsilon = \frac{1}{4}$, this computes the unit normal $(0,1)$ at $x = (0,0)$
- the Hessian matrix of $\varepsilon(u)$ is a large expression
- evaluated at $(su, s\varepsilon) = (0, \frac{s}{4}, \frac{s}{4})$ and letting $s \to 0$ yields $A = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$

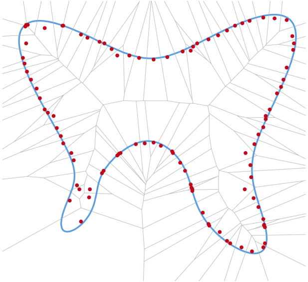
Voronoi Cells

Voronoi Cells

Definition: Let $X \subset \mathbb{R}^n$ and fix $y \in X$. The *Voronoi cell* of y is

$$\mathrm{Vor}_X(y) \;=\; \{u \in \mathbb{R}^n \mid y \in \arg\min_{x \in X} \|u - x\|\}.$$



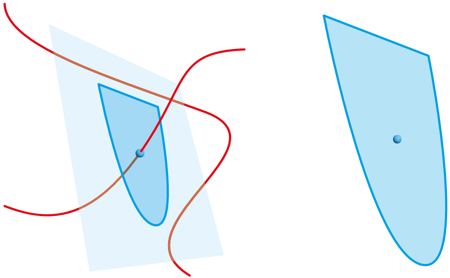


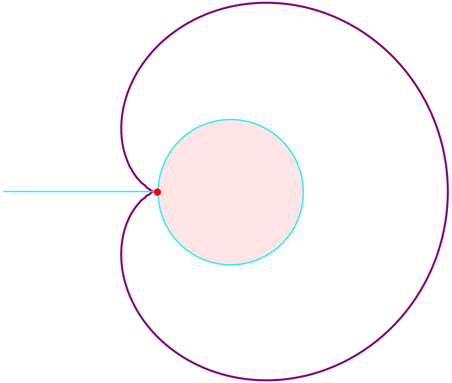
The union of the boundaries of the Voronoi cells is the medial axis.

Proposition: $X \subset \mathbb{R}^n$ algebraic variety, $y \in X$ is smooth. Then $\mathrm{Vor}_X(y)$ is a full-dimensional, convex, semialgebraic subset of the *affine normal space*

$$\begin{aligned} N_X(y) &= y + N_y X \\ &= \{u \mid u - y \perp T_y X\}. \end{aligned}$$

Voronoi Cells & Singularities



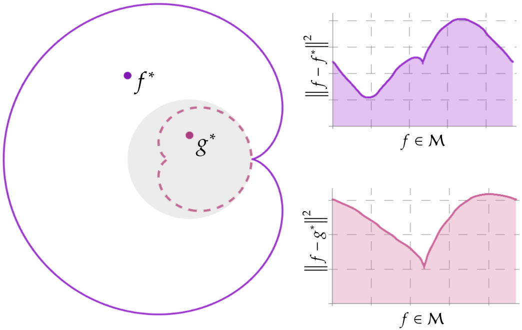


at a smooth point of a space curve

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

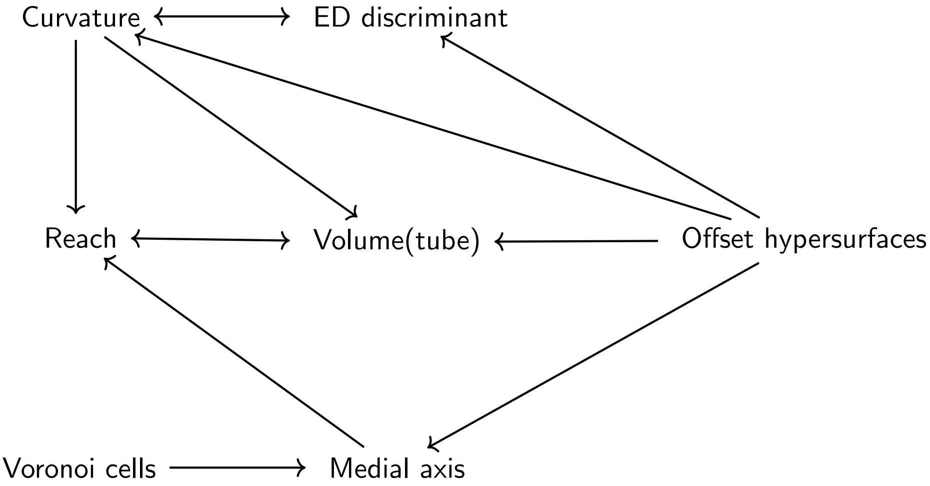
Singularities of neuromanifolds can cause implicit biases.

Voronoi Cells & ED discriminant



The number or type of critical points change when crossing the medial axis or the **ED discriminant**.

An Overview


Curvature
ED discriminant
Reach
Volume(tube)
Offset hypersurfaces
Voronoi cells
Medial axis