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 $\theta \in \Theta$ & $x \in X$)

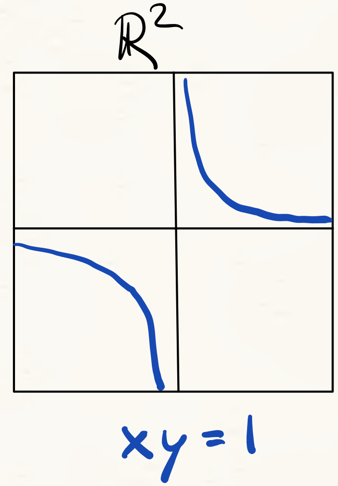
$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta, \cdot) \end{aligned}$$

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

A semialgebraic set!

describable by polynomial equations & inequalities

(Semi) Algebraic Sets

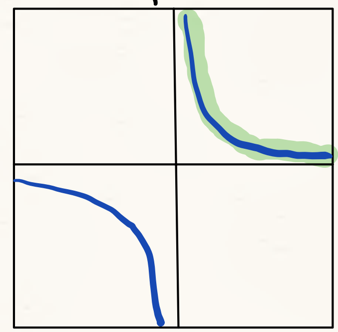


An **algebraic set / algebraic variety** in $\mathbb{R}^n$ is a solution set of a system of polynomial equations in $\mathbb{R}[x_1, \dots, x_n]$.

(Semi) Algebraic Sets

An algebraic set / algebraic variety in $\mathbb{R}^n$ is a solution set of a system of polynomial equations in $\mathbb{R}[x_1, \dots, x_n]$.

$\mathbb{R}^2$



$xy = 1$

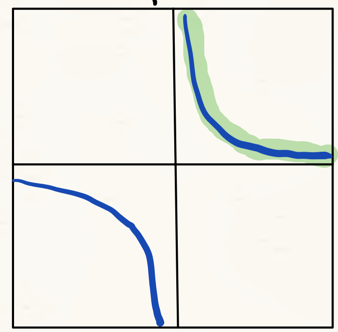
$x > 0$

A basic semialgebraic set in $\mathbb{R}^n$ is a solution set of a system of polynomial equations and polynomial inequalities in $\mathbb{R}[x_1, \dots, x_n]$.

(Semi) Algebraic Sets

An **algebraic set / algebraic variety** in $\mathbb{R}^n$ is a solution set of a system of polynomial equations in $\mathbb{R}[x_1,\dots,x_n]$.

$\mathbb{R}^2$

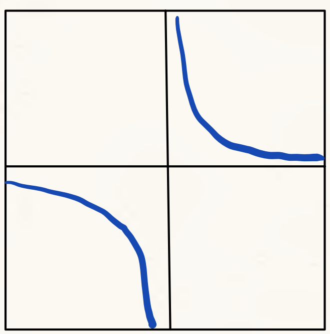


$xy = 1$

$x > 0$

A **basic semialgebraic set** in $\mathbb{R}^n$ is a solution set of a system of polynomial equations and polynomial inequalities in $\mathbb{R}[x_1,\dots,x_n]$.

A **semialgebraic set** is a finite union of basic semialgebraic sets.



$xy = 1$

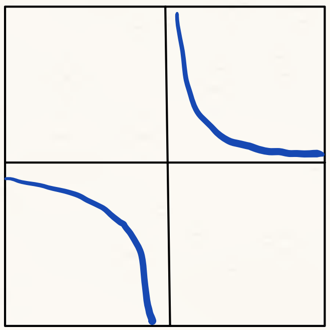
$(x,y) \longmapsto x$

$x<0 \;\cup\; x>0$

Tarski-Seidenberg Theorem

A **morphism between algebraic varieties** $\varphi : \underset{\mathbb{R}^n}{\overset{\cup\!\!\!\mid}{X}} \longrightarrow \underset{\mathbb{R}^m}{\overset{\cup\!\!\!\mid}{Y}}$ is a polynomial map, i.e., $\varphi = (\varphi_1, \dots, \varphi_m)$ & $\varphi_i \in \mathbb{R}[x_1, \dots, x_n]$.

Thm: $\varphi(X)$ is a semialgebraic set.



$xy = 1$

$$\xrightarrow{\qquad\qquad}$$

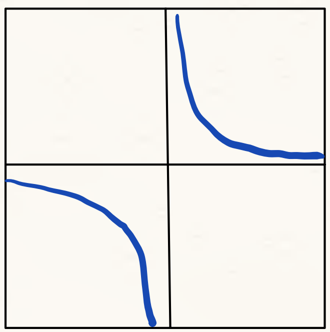
$(x,y) \longmapsto x$

$x<0 \;\cup\; x>0$

Tarski-Seidenberg Theorem

A **morphism between algebraic varieties** $\varphi : \overset{\mathbb{R}^n}{\overset{\cup\!\shortmid}{X}} \longrightarrow \overset{\mathbb{R}^m}{\overset{\cup\!\shortmid}{Y}}$ is a polynomial map, i.e., $\varphi = (\varphi_1, \dots, \varphi_m)$ & $\varphi_i \in \mathbb{R}[x_1, \dots, x_n]$.

Thm: $\varphi(X)$ is a semialgebraic set.



$xy = 1$

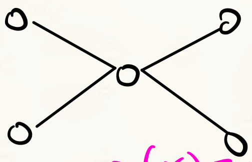
$$\xrightarrow{\qquad\qquad}$$
$$(x,y) \longmapsto x$$

$x<0 \;\cup\; x>0$

$\mu : \Theta \times X \longrightarrow Y$ polynomial in $x \in X$ $\rightsquigarrow$ $\mathcal{M} \subseteq V$ of finite dimension
& in $\theta \in \Theta$ $\rightsquigarrow$ $\mathcal{M}$ is semialgebraic

$$\Theta \longrightarrow \mathcal{M}$$
$$\theta \longmapsto \mu(\theta, \cdot)$$

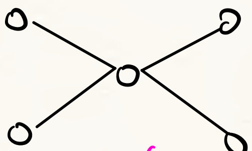
Linear MLPs:



$$\begin{bmatrix} c \\ d \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \implies \mathcal{M} = \left\{ W \in \mathbb{R}^{2\times 2} \mid \operatorname{rk}(W) \leq 1 \right\}$$

Is $\mathcal{M}$ a variety or just a semialgebraic set?

Linear MLPs:



$\sigma(x) = x$

$$\begin{bmatrix} c \\ d \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \Rightarrow \quad \mathcal{M} = \{ W \in \mathbb{R}^{2\times 2} \mid rk(W) \leq 1 \}$$

Is $\mathcal{M}$ a variety or just a semialgebraic set?

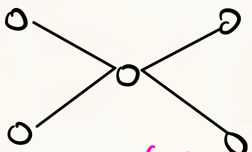
defined by $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = 0$

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

$$\Rightarrow \mathcal{M} = \{ W \in \mathbb{R}^{d_L \times d_0} \mid rk(W) \leq \min\{d_0, d_1, \dots, d_2\} \}$$

Is $\mathcal{M}$ a variety or just a semialgebraic set?

Linear MLPs:



$\sigma(x) = x$

$$\begin{bmatrix} c \\ d \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \Rightarrow \quad \mathcal{M} = \{ W \in \mathbb{R}^{2\times 2} \mid \mathrm{rk}(W) \leq 1 \}$$

Is $\mathcal{M}$ a variety or just a semialgebraic set?

defined by $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = 0$

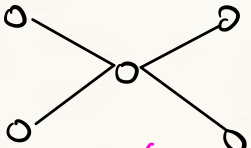
$\alpha_L \circ \dots \circ \alpha_2 \circ \alpha_1$, where $\alpha_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ linear

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

Is $\mathcal{M}$ a variety or just a semialgebraic set?

defined by the vanishing of all $(r+1)$-minors of W

Linear MLPs:



$\sigma(x) = x$

$$\begin{bmatrix} c \\ d \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \mathcal{M} = \{ W \in \mathbb{R}^{2 \times 2} \mid \mathrm{rk}(W) \leq 1 \}$$

Is $\mathcal{M}$ a variety or just a semialgebraic set?

defined by $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = 0$

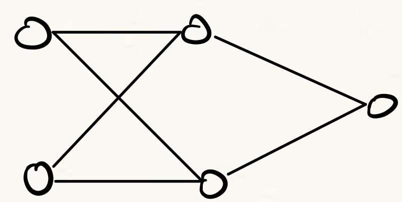
$\alpha_L \circ \dots \circ \alpha_2 \circ \alpha_1$, where $\alpha_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ linear

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

Is $\mathcal{M}$ a variety or just a semialgebraic set?

defined by the vanishing of all $(r+1)$-minors of W

Monomial MLPs: $\sigma(x) = x^3$



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

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

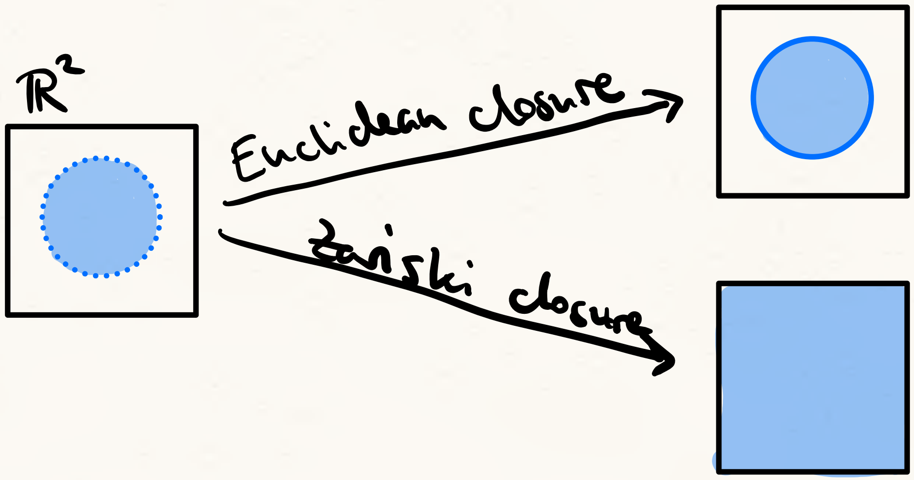
Is $\mathcal{M}$ a variety or just a semialgebraic set?

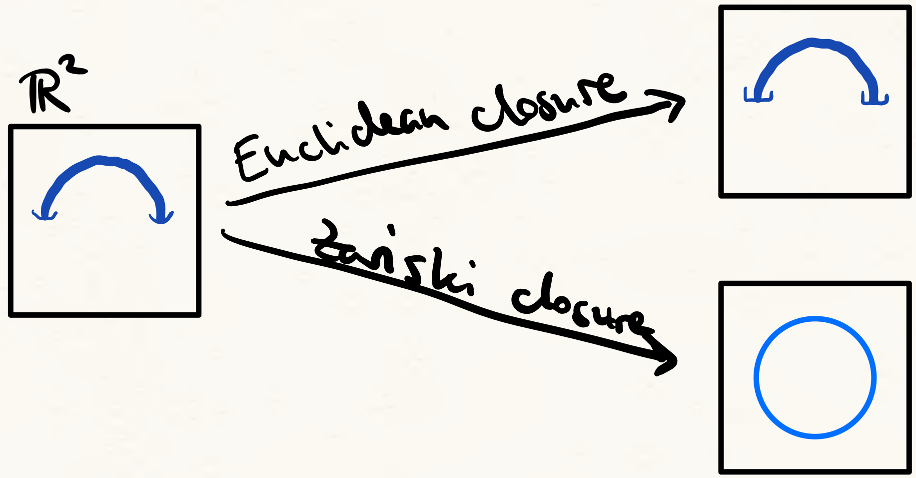
Intermezzo: Zariski topology & Dimension

Zariski topology on $\mathbb{R}^n$: closed sets are the algebraic varieties

The Zariski topology is coarser than the Euclidean topology:

- Zariski closed implies Euclidean closed
- Zariski open implies Euclidean open





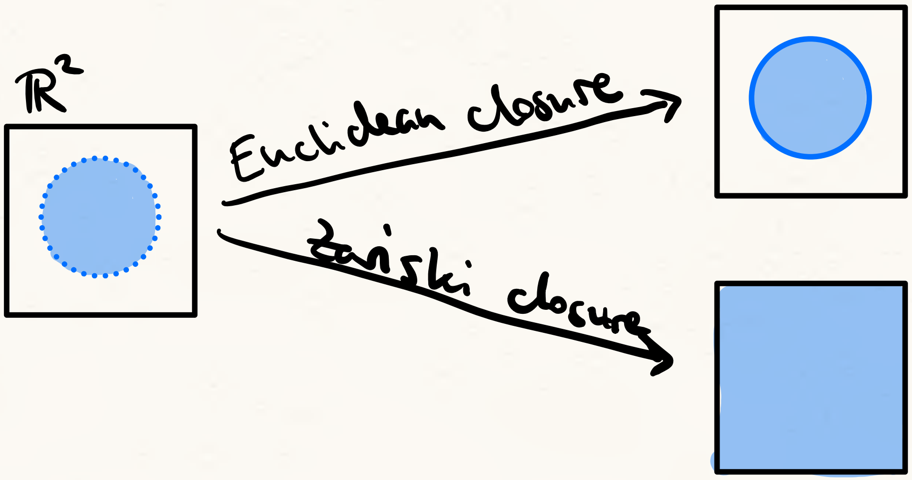
What are the Zariski closed sets in $\mathbb{R}^1$?

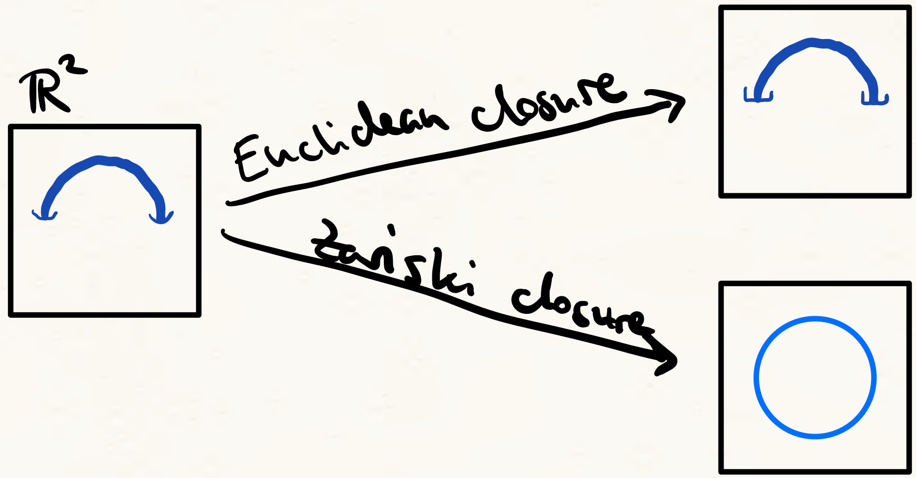
Intermezzo: Zariski topology & Dimension

Zariski topology on $\mathbb{R}^n$: closed sets are the algebraic varieties

The Zariski topology is coarser than the Euclidean topology:

- Zariski closed implies Euclidean closed
- Zariski open implies Euclidean open





What are the Zariski closed sets in $\mathbb{R}^1$?

What are the Zariski closed sets in $\mathbb{R}^2$?

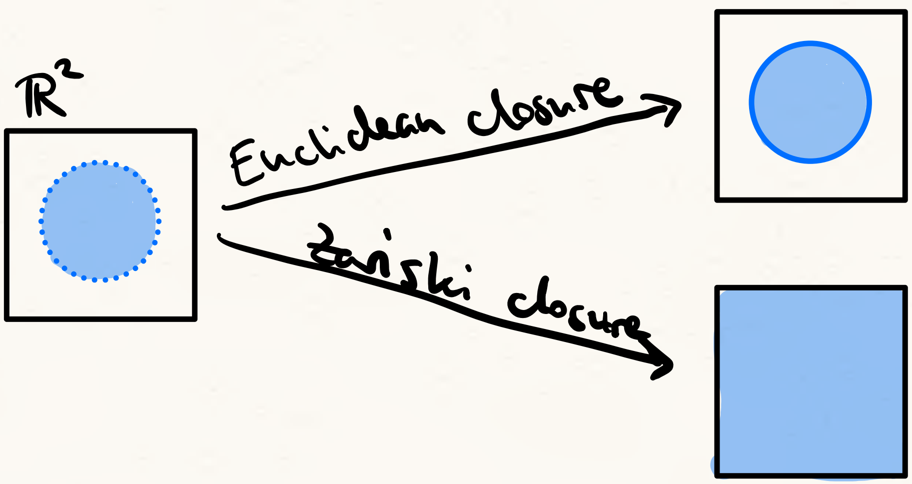
$\mathbb{R}^1$, finitely many points, $\emptyset$

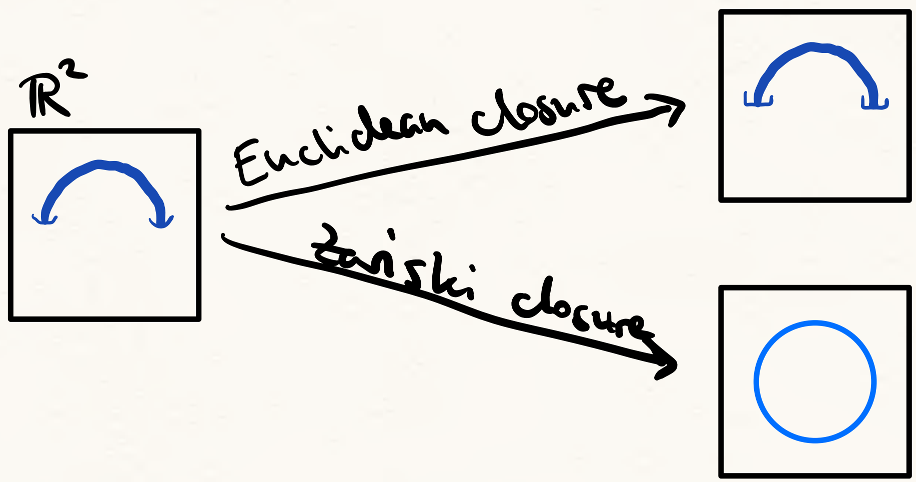
Intermezzo: Zariski topology & Dimension

Zariski topology on $\mathbb{R}^n$: closed sets are the algebraic varieties

The Zariski topology is coarser than the Euclidean topology:

- Zariski closed implies Euclidean closed
- Zariski open implies Euclidean open



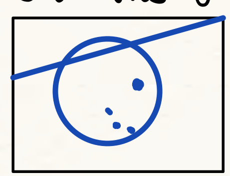


What are the Zariski closed sets in $\mathbb{R}^1$?

What are the Zariski closed sets in $\mathbb{R}^2$?

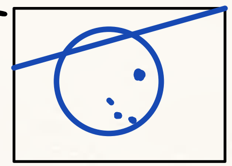
$\mathbb{R}^1$, finitely many points, $\emptyset$

$\mathbb{R}^2$, finite unions of algebraic curves & points, $\emptyset$



Intermezzo: Zariski topology & Dimension

A variety $X \subseteq \mathbb{R}^n$ is **irreducible** if it is not the union of 2 proper subvarieties,
i.e., there are no varieties $X_1, X_2 \subseteq \mathbb{R}^n$ s.t. $X = X_1 \cup X_2$
& $\emptyset \neq X_i \subsetneq X$

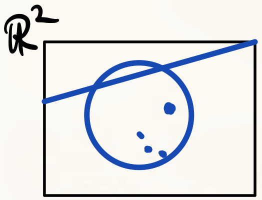
Ex.: $\mathbb{R}^2$  is reducible into 6 irreducible components

There is a subvariety $\Delta \subsetneq X$ such that $X \setminus \Delta$ is a smooth manifold of dimension k.
If X is irreducible, k is the same for all such Δ: k is the **dimension** of X.

$\dim(\bigcirc) = 1$, $\dim(\cdot) = 0$

Intermezzo: Zariski topology & Dimension

A variety $X \subseteq \mathbb{R}^n$ is **irreducible** if it is not the union of 2 proper subvarieties,
i.e., there are no varieties $X_1, X_2 \subseteq \mathbb{R}^n$ s.t. $X = X_1 \cup X_2$
& $\emptyset \neq X_i \subsetneq X$

Ex.: 

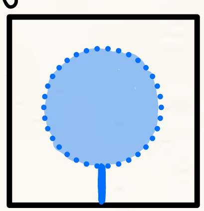
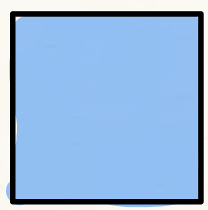
is reducible into 6 irreducible components

There is a subvariety $\Delta \subsetneq X$ such that $\underbrace{X \setminus \Delta}_{\text{Zariski open in } X}$ is a smooth manifold of dimension k.
If X is irreducible, k is the same for all such Δ: k is the **dimension** of X.

$\dim(\bigcirc) = 1$, $\dim(\cdot) = 0$

In general: $X = \bigcup_i X_i$ ← irreducible varieties
$\rightarrow \dim(X) := \max_i \dim(X_i)$

For a semialgebraic set $X \subseteq \mathbb{R}^n$, define $\dim(X) := \dim(\overline{X})$. ← Zariski closure

dim  = dim  = 2

Intermezzo: Zariski topology & Dimension

Let $\varphi : X \longrightarrow \mathbb{R}^m$ morphism between irreducible varieties.

Zariski closure

$$\Rightarrow Y := \overline{\varphi(X)} \text{ is irreducible}$$

Why?

Intermezzo: Zariski topology & Dimension

Let $\varphi : X \longrightarrow \mathbb{R}^m$ morphism between irreducible varieties.

Zariski closure

$\Rightarrow Y := \overline{\varphi(X)}$ is irreducible

Why?

almost everywhere in X /
for generic $x \in X$

Jacobian check: There is a subvariety $\Delta \subsetneq X$ such that $\underbrace{\text{for all } x \in X \setminus \Delta}$,
the rank of the Jacobi matrix $J_x(\varphi)$ equals $\dim(Y)$.

↳ practical test: ① choose random point $x \in X$
② compute $\operatorname{rank}(J_x(\varphi))$
③ repeat until confident

$\sigma(x) = x^3$

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

Compute dimension of $\mathcal{M} \subsetneq \mathbb{R}[x,y]_3$

Intermezzo: Zariski topology & Dimension

Let $\varphi : X \longrightarrow \mathbb{R}^m$ morphism between irreducible varieties.

Zariski closure

$\Rightarrow$ $Y := \overline{\varphi(X)}$ is irreducible **Why?**

Jacobian check: There is a subvariety $\Delta \subsetneq X$ such that $\overbrace{\text{for all } x \in X \setminus \Delta}^{\text{almost everywhere in } X \text{ / for generic } x \in X}$, the rank of the Jacobi matrix $J_x(\varphi)$ equals $\dim(Y)$.

↳ practical test: ① choose random point $x \in X$
② compute $\operatorname{rank}(J_x(\varphi))$
③ repeat until confident

$\sigma(x) = x^3$

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

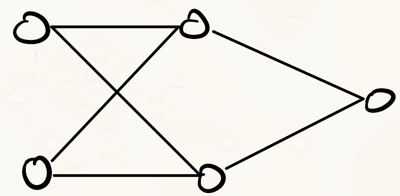
Compute dimension of $\mathcal{M} \subsetneq \mathbb{R}[x,y]_3$

Fact: $\operatorname{rank}(J_x(\varphi)) \leq \dim(Y)$ for <u>all</u> $x \in X$. **Why?**

↳ Finding a single point $x \in X$ with $\operatorname{rank}(J_x(\varphi)) = m$ is enough to show that $Y = \mathbb{R}^m$.

Monomial MLPs: $\sigma(x) = x^3$

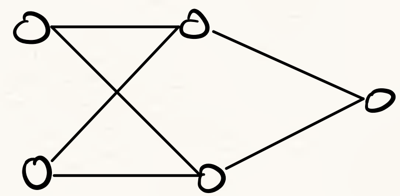
$$\begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$



$$\Rightarrow \underset{\dim = 4}{\mathcal{M} \subsetneq \mathbb{R}[x,y]_3}$$

Is $\mathcal{M}$ a variety or just a semialgebraic set?

Monomial MLPs: $\sigma(x) = x^3$



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

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

dim = 4

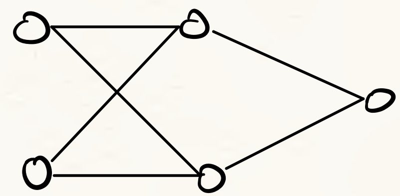
Is $\mathcal{M}$ a variety or just a semialgebraic set?

Is $\mathcal{M}$ Euclidean closed?

And why do we care?

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

Monomial MLPs: $\sigma(x) = x^3$



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

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

dim = 4

Is $\mathcal{M}$ a variety or just a semialgebraic set?

Is $\mathcal{M}$ Euclidean closed?

And why do we care?

parameter explosion!

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

No, e.g. $A = 1$, $B = 1$, $C = 0$, $D = 0$ $\notin \mathcal{M}$, but for $a = c = 1$, $b = 0$, $e = 1 - \frac{1}{3d}$, $f = \frac{1}{3d}$, we get

$$x^3 + x^2y + dxy^2 + \tfrac{1}{3}d^2y^3$$
$$\Big\downarrow {\scriptstyle d \to 0}$$
$$x^3 + x^2y$$

CNNs (convolutional neural nets)

on 1-dimensional signals without bias

$\alpha_L \circ \sigma \circ \dots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$, where each α_i is given by a "Toeplitz" matrix of the form

stride　　filter size

$$\begin{bmatrix} w_0 & w_1 & \cdots & w_s & \cdots & w_{k-1} & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & w_0 & w_1 & \cdots & w_{k-1} & 0 & \cdots & 0 \\ & & & & \ddots & & & \ddots & & \\ 0 & \cdots & & & & 0 & w_0 & w_1 & \cdots & w_{k-1} \end{bmatrix}$$

Proposition: If $\sigma(x) = x^r$, then CNN neuromanifold $\mathcal{M}$ is Euclidean closed.

CNNs (convolutional neural nets)

on 1-dimensional signals without bias

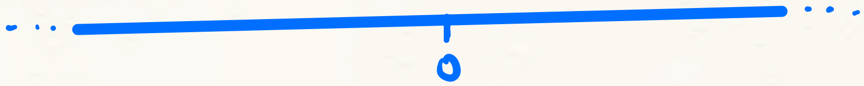
$\alpha_L \circ \sigma \circ \dots \circ \sigma \circ \alpha_2 \circ \sigma \circ \alpha_1$, where each α_i is given by a "Toeplitz" matrix of the form

$$\begin{array}{c}
\hspace{5em}\text{stride} \hspace{6em} \text{filter size}\\
\begin{bmatrix} w_0 & w_1 & \cdots & w_s & \cdots & w_{k-1} & 0 & \cdots & & & 0 \\ 0 & \cdots & 0 & w_0 & w_1 & \cdots & w_{k-1} & 0 & \cdots & & 0 \\ & & & & \ddots & & & \ddots & & & \\ 0 & \cdots & & & & & 0 & w_0 & w_1 & \cdots & w_{k-1} \end{bmatrix}
\end{array}$$

Proposition: If $\sigma(x) = x^r$, then CNN neuromanifold $\mathcal{M}$ is Euclidean closed.

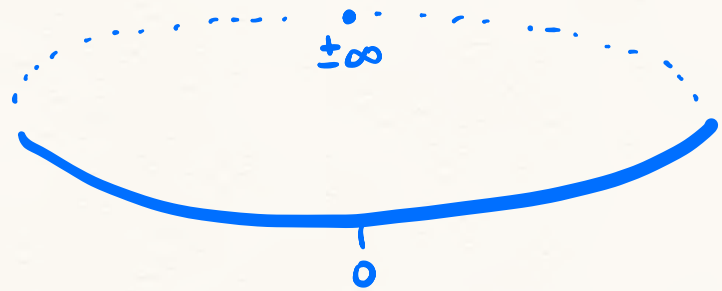
↳ **Reason:**
$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta, \cdot) : X \to Y \end{aligned}$$
is a projective morphism

Issue: Affine space is not compact



affine line $\mathbb{R}^1$

Projective space is compact ☺

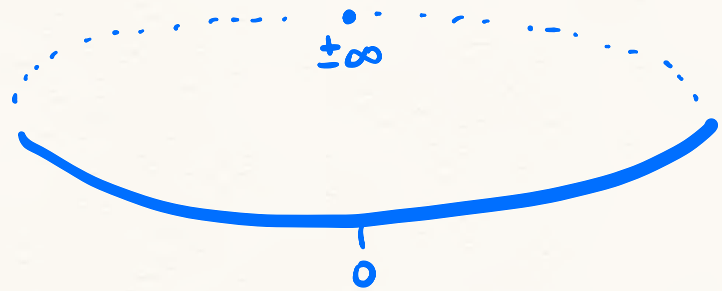


projective line $\mathbb{P}^1_{\mathbb{R}} = \mathbb{R}^1 \cup \{\pm\infty\}$

Issue: Affine space is not compact

affine line $\mathbb{R}^1$

Projective space is compact

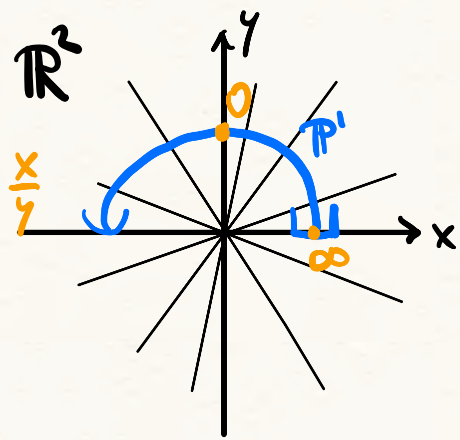


projective line $\mathbb{P}^1_{\mathbb{R}} = \mathbb{R}^1 \cup \{\pm\infty\}$

standard construction:

$\mathbb{P}^n_{\mathbb{R}} = (\mathbb{R}^{n+1} \setminus \{0\})/\sim$, where

$u \sim v \;:\Leftrightarrow\; \exists \lambda \in \mathbb{R}\setminus\{0\}: u = \lambda v$

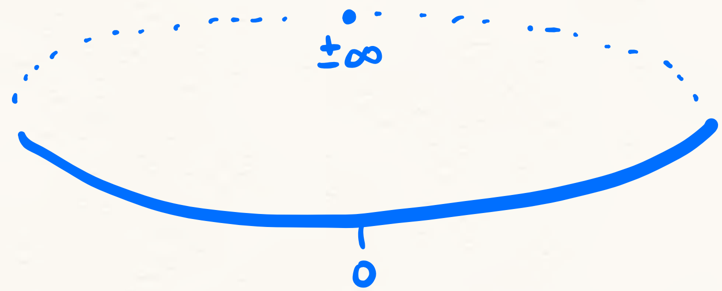


$\mathbb{P}^n_{\mathbb{R}}$ = set of lines in $\mathbb{R}^{n+1}$ through origin

Issue: Affine space is not compact

affine line $\mathbb{R}^1$

Projective space is compact ☺



projective line $\mathbb{P}^1_{\mathbb{R}} = \mathbb{R}^1 \cup \{\pm\infty\}$

compact

$\mathbb{P}^n_{\mathbb{R}}$ is the sphere in $\mathbb{R}^{n+1}$ after identifying antipodal points

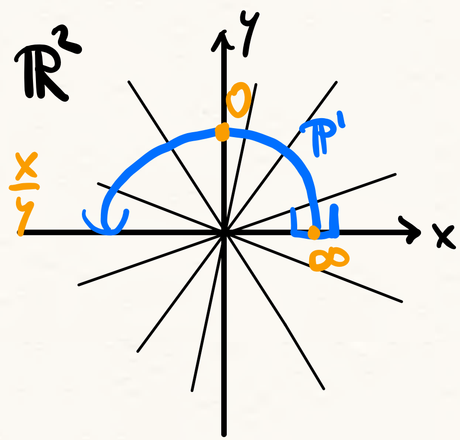
$\Rightarrow$ $\mathbb{P}^n_{\mathbb{R}}$ is compact in the quotient topology of the Euclidean topology on $\mathbb{R}^{n+1}$

standard construction:

$\mathbb{P}^n_{\mathbb{R}} = (\mathbb{R}^{n+1} \setminus \{0\})/\sim$, where

$u \sim v \;:\Leftrightarrow\; \exists \lambda \in \mathbb{R}\setminus\{0\} : u = \lambda v$

$\mathbb{R}^2$



$\mathbb{P}^n_{\mathbb{R}}$ = set of lines in $\mathbb{R}^{n+1}$ through origin

Standard fact from topology:

$\varphi : T_1 \to T_2$ continuous map between topological spaces,
T_1 compact
$\Longrightarrow$ $\mathrm{im}(\varphi)$ is compact

Standard fact from topology:

$$\boxed{\begin{array}{l} \varphi : T_1 \to T_2 \text{ continuous map between topological spaces,} \\ \quad T_1 \text{ compact} \\ \implies \mathrm{im}(\varphi) \text{ is compact} \end{array}}$$

not compact ☹

$\sigma(x) = x^3$

$$\varphi_{MLP} : \underbrace{\mathbb{R}^{2\times 2} \times \mathbb{R}^{1\times 2}} \longrightarrow \mathbb{R}[x,y]_3$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} e & f \end{bmatrix} \right) \longmapsto \begin{bmatrix} e & f \end{bmatrix} \sigma \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) \quad \Big\} \text{ continuous ☺}$$

Can we projectivize φ_{MLP}?

$$\varphi_{MLP}\left(\alpha\begin{bmatrix} a & b \\ c & d\end{bmatrix}, \beta\begin{bmatrix} e & f\end{bmatrix}\right) = \beta\begin{bmatrix} e & f\end{bmatrix}\sigma\left(\alpha\begin{bmatrix} a & b \\ c & d\end{bmatrix}\begin{bmatrix} x \\ y\end{bmatrix}\right) = \alpha^{3}\beta\,\varphi_{MLP}\left(\begin{bmatrix} a & b \\ c & d\end{bmatrix}, \begin{bmatrix} e & f\end{bmatrix}\right)$$

so φ_{MLP} is well-behaved under scaling the factors and it makes sense to consider

$$\tilde{\varphi}_{MLP} : \mathbb{P}^3_{\mathbb{R}} \times \mathbb{P}^1_{\mathbb{R}} \longrightarrow \mathbb{P}^3_{\mathbb{R}}$$

$$\left(\left[\begin{bmatrix} a & b \\ c & d\end{bmatrix}\right], \left[\begin{bmatrix} e & f\end{bmatrix}\right]\right) \longmapsto \left[\begin{array}{l}(a^3e + c^3f)x^3 + 3(a^2be + c^2df)x^2y \\ + 3(ab^2e + cd^2f)xy^2 + (b^3e + d^3f)y^3\end{array}\right]$$

defined up to scaling!

BUT ...

what is the issue?

$$\varphi_{MLP}\left(\alpha\begin{bmatrix}a&b\\c&d\end{bmatrix},\ \beta[e\ f]\right) = \beta[e\ f]\,\sigma\left(\alpha\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}\right) = \alpha^3\beta\,\varphi_{MLP}\left(\begin{bmatrix}a&b\\c&d\end{bmatrix},\ [e\ f]\right)$$

so φ_{MLP} is well-behaved under scaling the factors and it makes sense to consider

$$\tilde{\varphi}_{MLP} : \mathbb{P}^3_{\mathbb{R}} \times \mathbb{P}^1_{\mathbb{R}} \longrightarrow \mathbb{P}^3_{\mathbb{R}}$$

$$\left(\left[\begin{bmatrix}a&b\\c&d\end{bmatrix}\right],\ \big[[e\ f]\big]\right) \longmapsto \left[\begin{array}{l}(a^3e+c^3f)x^3 + 3(a^2be+c^2df)x^2y \\ \quad + 3(ab^2e+cd^2f)xy^2 + (b^3e+d^3f)y^3\end{array}\right]$$

defined up to scaling!

BUT $\tilde{\varphi}_{MLP}$ is not defined everywhere!

E.g. $\varphi_{MLP}\left(\begin{bmatrix}1&1\\1&1\end{bmatrix},\ [1\ -1]\right) = 0 \notin \mathbb{P}^3_{\mathbb{R}}$

cannot apply topology fact!

Compute all points in $\mathbb{P}^2_{\mathbb{R}} \times \mathbb{P}^1_{\mathbb{R}}$, where $\tilde{\varphi}_{MLP}$ is not defined.

This is called the base locus of $\tilde{\varphi}_{MLP}$.

Standard fact from topology:

$\varphi : T_1 \to T_2$ continuous map between topological spaces,
T_1 compact
$\implies$ $\mathrm{im}(\varphi)$ is compact

$$\varphi_{CNN} : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1, \dots, x_5]_3$$
$$\big((a,b,c),(e,f)\big) \longmapsto [e\ f]\ \sigma\left(\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix}\right)$$

$\sigma(x) = x^3$

can be projectivized: $\tilde{\varphi}_{CNN} : \mathbb{P}^2_{\mathbb{R}} \times \mathbb{P}^1_{\mathbb{R}} \longrightarrow \mathbb{P}(\mathbb{R}[x_1, \dots, x_5]_3)$

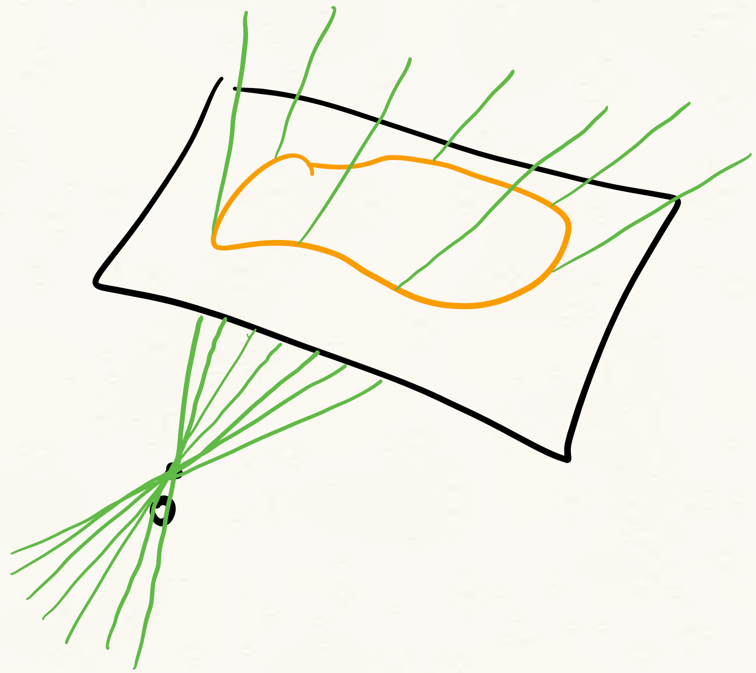
$$\big([(a,b,c)],[(e,f)]\big) \longmapsto \left[[e\ f]\ \sigma\left(\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix} x\right)\right]$$

Compute: $\tilde{\varphi}_{LCN}$ is defined everywhere, i.e., its base locus is empty. $\Big\}$ can apply topology fact

Hence: $\mathrm{im}(\tilde{\varphi}_{CNN}) \subseteq \mathbb{P}(\mathbb{R}[x_1, \ldots, x_5]_3)$ is compact (in Euclidean quotient top.).

Since $\mathrm{im}(\varphi_{CNN}) \subseteq \mathbb{R}[x_1, \ldots, x_5]_3$ is the affine cone over $\mathrm{im}(\tilde{\varphi}_{CNN})$, it is closed (in Euclidean topology).

replace each projective point by the affine line it represents



Proposition: If $\sigma(x) = x^r$, then CNN neuromanifold $\mathcal{M}$ is Euclidean closed.

↳ **Reason:**
$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta, \cdot) : X \to Y \end{aligned}$$
is a projective morphism

Proposition: If $\sigma(x) = x^r$, then CNN neuromanifold $\mathcal{M}$ is Euclidean closed.

↳ **Reason:**
$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta, \cdot) : X \to Y \end{aligned}$$
is a projective morphism

Example: A linear CNN ($\sigma(x) = x$):

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1, \dots, x_5]_1 \cong \mathbb{R}^5 \qquad \mathcal{M} := \mathrm{im}(\varphi)$$
$$((a,b,c),(e,f)) \longmapsto [e\ f] \begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix},$$

① Compute that $\dim(\mathcal{M}) = 4$. ⇝ $\overline{\mathcal{M}}$ is a hypersurface and thus defined by a single equation.

Zariski closure

② Find that equation!

③ Check that base locus of projectivization $\tilde{\varphi} : \mathbb{P}^2_{\mathbb{R}} \times \mathbb{P}^1_{\mathbb{R}} \to \mathbb{P}^4_{\mathbb{R}}$ is empty.

④ Find a point in $\overline{\mathcal{M}} \setminus \mathcal{M}$.

Proposition: If $\sigma(x) = x^r$, then CNN neuromanifold $\mathcal{M}$ is Euclidean closed.

↳ **Reason:**
$$\begin{aligned} \Theta &\longrightarrow \mathcal{M} \\ \theta &\longmapsto \mu(\theta,\cdot): X \to Y \end{aligned}$$
is a projective morphism

Example: A linear CNN ($\sigma(x)=x$):

$$\begin{aligned} \varphi: \mathbb{R}^3 \times \mathbb{R}^2 &\longrightarrow \mathbb{R}[x_1,\ldots,x_5]_1 \cong \mathbb{R}^5 \\ ((a,b,c),(e,f)) &\longmapsto [e\ f]\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} = Ax_1+Bx_2+Cx_3+Dx_4+Ex_5 \end{aligned}$$

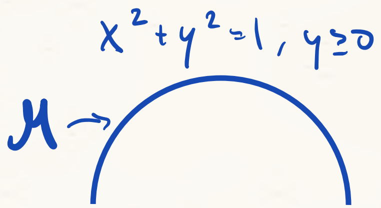
$\mathcal{M} := \mathrm{im}(\varphi)$

Zariski closure

1. Compute that $\dim(\mathcal{M}) = 4$. ⇝ $\overline{\mathcal{M}}$ is a hypersurface and thus defined by a single equation.
2. Find that equation! $AD^2+B^2E-BCD=0$
3. Check that base locus of projectivization $\tilde{\varphi}: \mathbb{P}^2_{\mathbb{R}} \times \mathbb{P}^1_{\mathbb{R}} \to \mathbb{P}^4_{\mathbb{R}}$ is empty.
4. Find a point in $\overline{\mathcal{M}} \setminus \mathcal{M}$. $B=C=D=0,\ A=E=1$

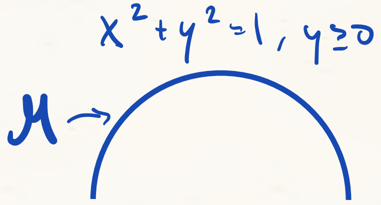
Euclidean boundary?

What is the Euclidean boundary of $\mathcal{M}$? (And why do we care?)



Euclidean boundary?

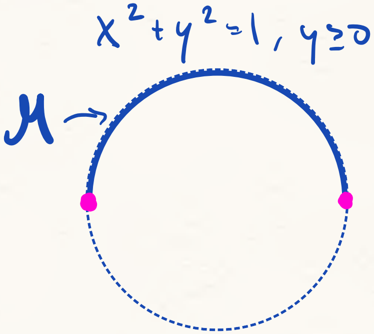
What is the Euclidean boundary of M? (And why do we care?)



Euclidean boundary of $M \subseteq \mathbb{R}^2$ is M,
so its interior is empty

Euclidean boundary?

What is the Euclidean boundary of $\mathcal{M}$? (And why do we care?)



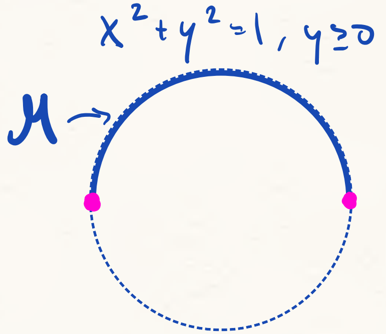
Euclidean boundary of $\mathcal{M} \subseteq \mathbb{R}^2$ is $\mathcal{M}$,
so its interior is empty

Its boundary in the Euclidean topology on $\overline{\mathcal{M}}$ (Zariski closure) (inherited from $\mathbb{R}^2$)
is 2 points.

In general: We write $\partial \mathcal{M}$ for the relative Euclidean boundary of $\mathcal{M}$ inside $\overline{\mathcal{M}}$ (Zariski closure).

Euclidean boundary?

What is the Euclidean boundary of $\mathcal{M}$? (And why do we care?)



Euclidean boundary of $\mathcal{M} \subseteq \mathbb{R}^2$ is $\mathcal{M}$,
so its interior is empty

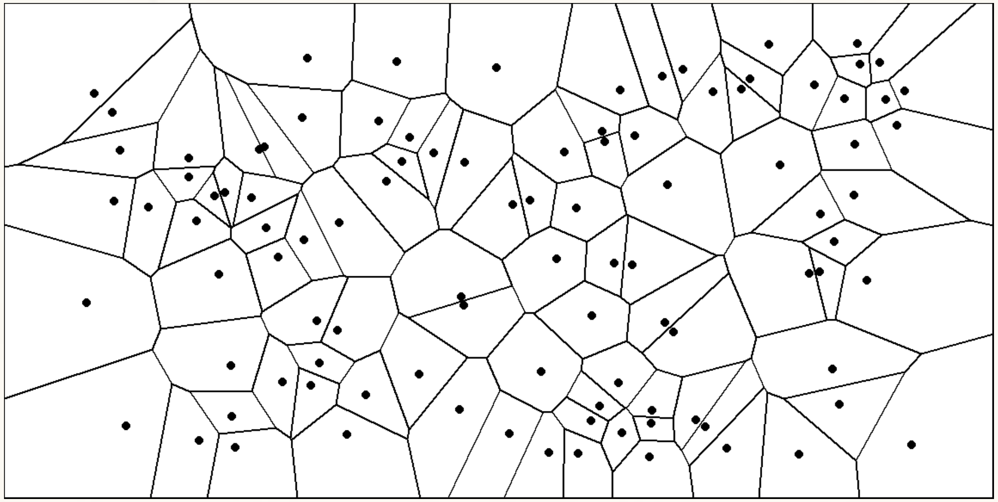
Its boundary in the Euclidean topology on $\overline{\mathcal{M}}$ (inherited from $\mathbb{R}^2$)
is 2 points.

Zariski closure

In general: We write $\partial\mathcal{M}$ for the relative Euclidean boundary of $\mathcal{M}$ inside $\overline{\mathcal{M}}$.

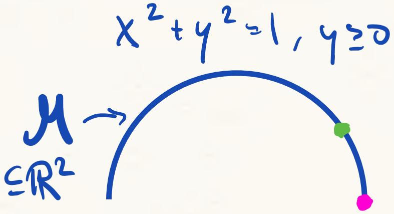
Zariski closure

Voronoi cells



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

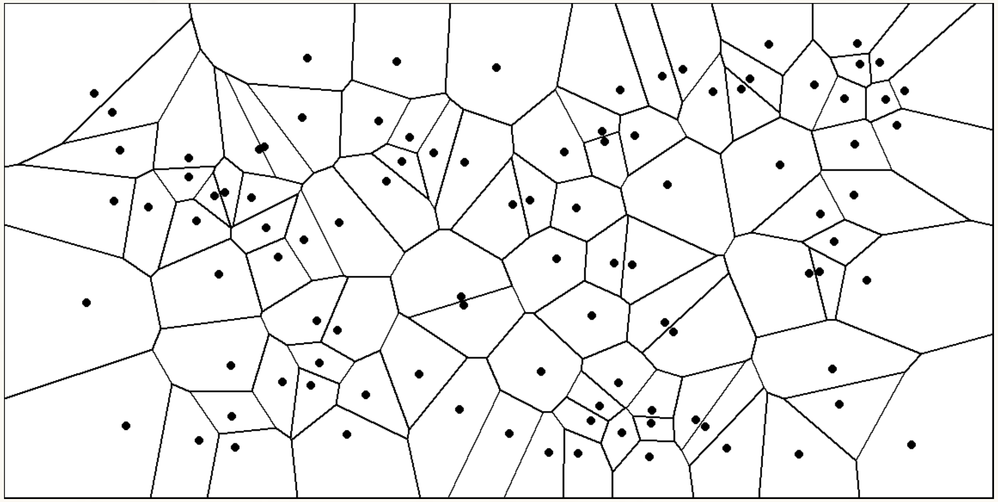
$$\mathrm{Vor}_S(p) := \{ u \in \mathbb{R}^n \mid \forall q \in S,\ q \neq p : \|p-u\|_2 < \|q-u\|_2 \}$$



What is the Voronoi cell at ● ?

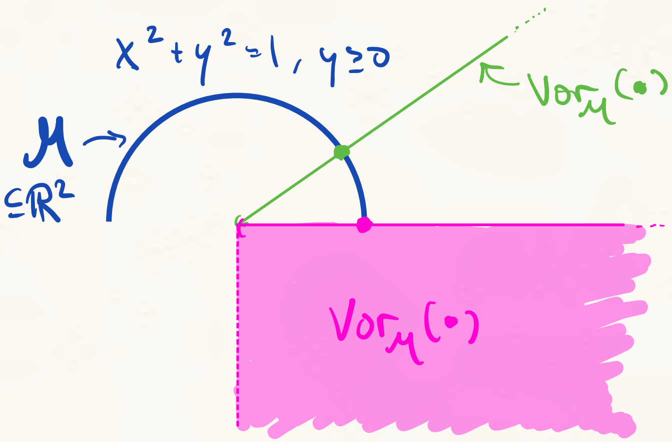
What is the Voronoi cell at ● ?

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

Relative Euclidean boundary $\partial \mathcal{M}$

$\underline{\partial \mathcal{M} \setminus \mathcal{M}}$

Causes parameter explosion

$\underline{\partial \mathcal{M} \cap \mathcal{M}}$

can cause implicit bias

How to compute $\partial \mathcal{M}$?

Relative Euclidean boundary $\partial \mathcal{M}$

$\underline{\partial \mathcal{M} \setminus \mathcal{M}}$

causes parameter explosion

$\underline{\partial \mathcal{M} \cap \mathcal{M}}$

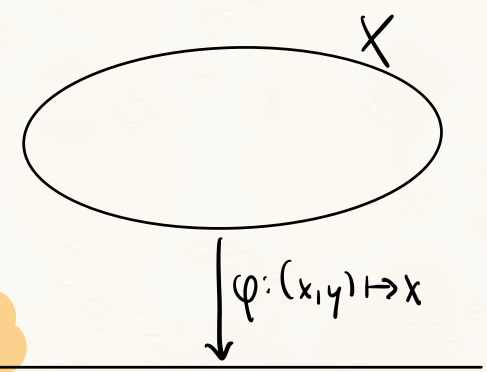
can cause implicit bias

How to compute $\partial \mathcal{M}$?

Let $\varphi : X \longrightarrow \mathbb{R}^m$ morphism between irreducible varieties, $Y := \overline{\varphi(X)}$. Zariski closure

The ranification locus of φ is $\mathrm{Ram}(\varphi) := \{x \in X \mid \mathrm{rank}\ J_x(\varphi) < \dim Y\}$.

The branch locus of φ is $\mathrm{Br}(\varphi) := \varphi(\mathrm{Ram}(\varphi))$.



What are $\mathrm{Ram}(\varphi)$ & $\mathrm{Br}(\varphi)$ in this example?

Relative Euclidean boundary $\partial\mathcal{M}$

$\underline{\partial\mathcal{M} \setminus \mathcal{M}}$ causes parameter explosion

$\underline{\partial\mathcal{M} \cap \mathcal{M}}$ can cause implicit bias

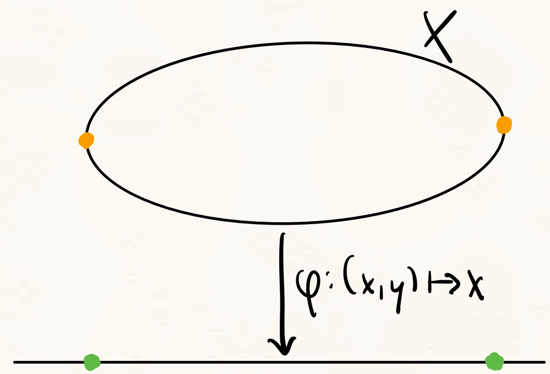
How to compute $\partial\mathcal{M}$?

Let $\varphi : X \longrightarrow \mathbb{R}^m$ morphism between irreducible varieties, $Y := \overline{\varphi(X)}$.

Zariski closure

The ramification locus of φ is $\mathrm{Ram}(\varphi) := \{x \in X \mid \mathrm{rank}\, J_x(\varphi) < \dim Y\}$.

The branch locus of φ is $\mathrm{Br}(\varphi) := \varphi(\mathrm{Ram}(\varphi))$.



Lemma: $\partial\, \mathrm{im}(\varphi) \cap \mathrm{im}(\varphi) \subseteq \mathrm{Br}(\varphi)$

↳ Why?

↳ Is this an equality?

Relative Euclidean boundary $\partial\mathcal{M}$

$\partial\mathcal{M} \setminus \mathcal{M}$ — causes parameter explosion

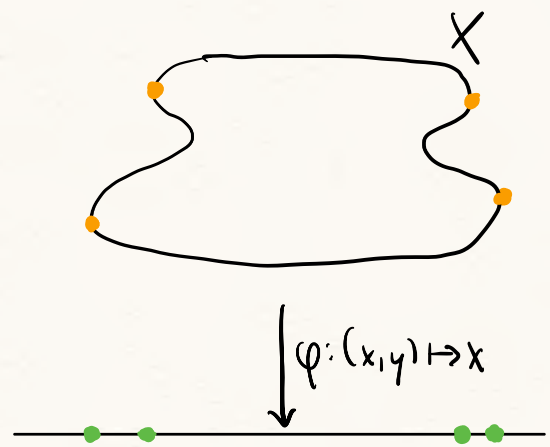
$\partial\mathcal{M} \cap \mathcal{M}$ — can cause implicit bias

How to compute $\partial\mathcal{M}$?

Let $\varphi : X \longrightarrow \mathbb{R}^m$ morphism between irreducible varieties, $Y := \overline{\varphi(X)}$. ← Zariski closure

The ramification locus of φ is $\mathrm{Ram}(\varphi) := \{x \in X \mid \operatorname{rank} J_x(\varphi) < \dim Y\}$.

The branch locus of φ is $\mathrm{Br}(\varphi) := \varphi(\mathrm{Ram}(\varphi))$.



Lemma: $\partial \operatorname{im}(\varphi) \cap \operatorname{im}(\varphi) \subseteq \mathrm{Br}(\varphi)$

↳ Why? essentially inverse function theorem...

↳ Is this an equality? **not in general**

Compute branch locus (& $\partial\mathcal{M}$) in the following examples:

↖ challenge!

① A linear CNN ($\sigma(x)=x$):

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1,\dots,x_5]_1 \cong \mathbb{R}^5$$

$$((a,b,c),(e,f)) \longmapsto [e\ f]\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix}$$

② A monomial CNN:

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1,\dots,x_5]_3$$

$$((a,b,c),(e,f)) \longmapsto [e\ f]\ \sigma\left(\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix}\right)$$

$\sigma(x) = x^3$

③ A monomial MLP:

$\sigma(x)=x^3$

$$\varphi : \mathbb{R}^{2\times 2} \times \mathbb{R}^{1\times 2} \longrightarrow \mathbb{R}[x,y]_3$$

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

Compute branch locus (& $\partial\mathcal{M}$) in the following examples:

↖ challenge!

① A linear CNN ($\sigma(x)=x$):

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1,\dots,x_5]_1 \cong \mathbb{R}^5$$

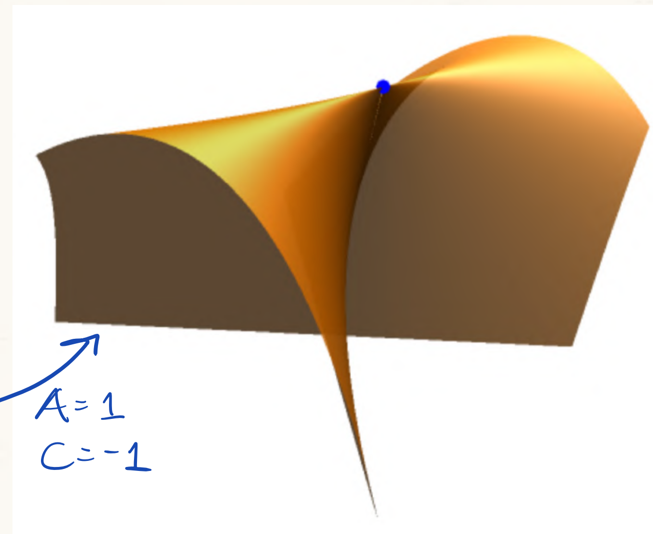
$$((a,b,c),(e,f)) \longmapsto \begin{bmatrix} e & f \end{bmatrix} \begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} = \underbrace{ae}_{A}x_1 + \underbrace{be}_{B}x_2 + \underbrace{(ce+af)}_{C}x_3 + \underbrace{bf}_{D}x_4 + \underbrace{cf}_{E}x_5$$

$\mathrm{Ram}(\varphi) = \{e=f=0\} \cup \{b = 0 = \det\begin{bmatrix} a & c \\ e & f \end{bmatrix}\}$

$\overline{\mathrm{Br}(\varphi)} = \{0 = B = D = C^2-4AE\}$

$\mathcal{M} \subseteq \{AD^2+B^2E-BCD=0,\ C^2-4AE \geq 0\}$

↖ [in fact equality!]



For any (A,C,E) with $C^2=4AE$, find sequence $(A_\varepsilon,C_\varepsilon,E_\varepsilon) \longrightarrow (A,C,E)$

such that $B_\varepsilon=0$, $D_\varepsilon=0$, $\rightsquigarrow \in \overline{\mathcal{M}}$

$0 > C_\varepsilon^2 - 4A_\varepsilon E_\varepsilon \rightsquigarrow \notin \mathcal{M}$

$\Rightarrow \partial\mathcal{M} = \mathrm{Br}(\varphi) = \overline{\mathrm{Br}(\varphi)}$

Compute branch locus (& $\partial \mathcal{M}$) in the following examples:

$\hookleftarrow$ challenge!

② A monomial CNN:

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1, \dots, x_5]_3$$

$$((a,b,c),(e,f)) \longmapsto [e \ f] \, \sigma\left(\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} \right)$$

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

$\operatorname{Ram}(\varphi) = \{e = f = 0\} \cup \{a = b = c = 0\}$

$\operatorname{Br}(\varphi) = \{0\}$

$$\Rightarrow \partial \mathcal{M} \underset{\substack{\uparrow \\ \text{Euclidean} \\ \text{closed}}}{=} \partial \mathcal{M} \cap \mathcal{M} \subseteq \{0\}$$

Since $\mathcal{M}$ is scaling invariant (i.e., a projective variety where $\operatorname{Br}(\tilde{\varphi}) = \emptyset$), $\mathcal{M} = \overline{\mathcal{M}}$ & $\partial \mathcal{M} = \emptyset$

Compute branch locus (& $\partial\mathcal{M}$) in the following examples:

challenge!

② A monomial CNN:

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^2 \longrightarrow \mathbb{R}[x_1, \dots, x_5]_3$$

$$((a,b,c),(e,f)) \longmapsto [e\ f]\ \sigma\left(\begin{bmatrix} a & b & c & 0 & 0 \\ 0 & 0 & a & b & c \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix}\right)$$

$\sigma(x) = x^3$

$\mathrm{Ram}(\varphi) = \{e=f=0\} \cup \{a=b=c=0\}$
$\mathrm{Br}(\varphi) = \{0\}$

$\Rightarrow \partial\mathcal{M} \underset{\substack{\uparrow \\ \text{Euclidean closed}}}{=} \partial\mathcal{M} \cap \mathcal{M} \subseteq \{0\}$

Theorem [Shahverdi, Marchetti, K.]:
For every monomial CNN with $\sigma(x) = x^r$, $r > 1$, without bias vectors:
$\mathrm{Ram}(\tilde{\varphi}) = \emptyset$ & $\mathcal{M}$ is Zariski closed.

Since $\mathcal{M}$ is scaling invariant (i.e., a projective variety where $\mathrm{Br}(\tilde{\varphi}) = \emptyset$),
$\mathcal{M} = \overline{\mathcal{M}}$ & $\partial\mathcal{M} = \emptyset$

Compute branch locus (& ∂M) in the following examples:

↖ challenge!

(3) A monomial MLP:

$\sigma(x) = x^3$

$$\varphi: \mathbb{R}^{2\times 2} \times \mathbb{R}^{1\times 2} \longrightarrow \mathbb{R}[x,y]_3$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} e & f \end{bmatrix} \right) \longmapsto \begin{bmatrix} e & f \end{bmatrix} \sigma \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) = e(ax+by)^3 + f(cx+dy)^3$$

$$\mathrm{Ram}(\varphi) = \{e=0\} \cup \{f=0\} = \{\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = 0\}$$

$$\longrightarrow \mathrm{Br}(\varphi) = \{(\alpha x + \beta y)^3 \mid \alpha, \beta \in \mathbb{R}\}$$

Compute branch locus (& $\partial\mathcal{M}$) in the following examples:

← challenge!

③ A monomial MLP:

$\sigma(x) = x^3$

$$\varphi : \mathbb{R}^{2\times 2} \times \mathbb{R}^{1\times 2} \longrightarrow \mathbb{R}[x,y]_3$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} e & f \end{bmatrix} \right) \longmapsto \begin{bmatrix} e & f \end{bmatrix} \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) = e(ax+by)^3 + f(cx+dy)^3$$

$$= Ax^3 + Bx^2y + Cxy^2 + Dy^3$$

$\mathrm{Ram}(\varphi) = \{e=0\} \cup \{f=0\} = \{\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = 0\}$

$\Rightarrow \mathrm{Br}(\varphi) = \{(\alpha x + \beta y)^3 \mid \alpha, \beta \in \mathbb{R}\}$

discriminant of

$\Rightarrow \mathcal{M} \subseteq \{B^2C^2 - 4AC^3 - 4B^3D - 27A^2D^2 + 18ABCD \leq 0\}$

Recall: $\overline{\mathcal{M}} = \mathbb{R}[x,y]_3$

For $(\alpha x + \beta y)^3$, consider sequence $(\alpha x + \beta(y+\varepsilon)) \cdot (\alpha x + \beta y) \cdot (\alpha x + \beta(y-\varepsilon)) \notin \mathcal{M}$

$\Rightarrow \partial\mathcal{M} = \mathrm{Br}(\varphi)$

Compute branch locus (& $\partial\mathcal{M}$) in the following examples:

challenge!

③ A monomial MLP:

$\sigma(x) = x^3$

$$\varphi : \mathbb{R}^{2\times 2} \times \mathbb{R}^{1\times 2} \longrightarrow \mathbb{R}[x,y]_3$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, [e\ f] \right) \longmapsto [e\ f]\, \sigma\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) = e(ax+by)^3 + f(cx+dy)^3$$

$$= Ax^3 + Bx^2y + Cxy^2 + Dy^3$$

$\mathrm{Ram}(\varphi) = \{e=0\} \cup \{f=0\} = \{\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = 0\}$

$\rightarrow \mathrm{Br}(\varphi) = \{(\alpha x + \beta y)^3 \mid \alpha, \beta \in \mathbb{R}\}$

discriminant of

$\Rightarrow \mathcal{M} \subseteq \{B^2C^2 - 4AC^3 - 4B^3D - 27A^2D^2 + 18ABCD \leq 0\}$ = Euclidean closure of $\mathcal{M}$

Recall: $\overline{\mathcal{M}} = \mathbb{R}[x,y]_3$

For $(\alpha x + \beta y)^3$, consider sequence $(\alpha x + \beta(y+\varepsilon)) \cdot (\alpha x + \beta y) \cdot (\alpha x + \beta(y-\varepsilon)) \notin \mathcal{M}$

$\Rightarrow \partial\mathcal{M} = \mathrm{Br}(\varphi)$

<u>not</u> Euclidean closed

Challenge: Compute $\mathcal{M}$!

Overview (no bias vectors)

	linear $\sigma(x) = x$	monomial $\sigma(x) = x^r,\ r > 1$
MLPs	$\mathcal{M}$ Zariski closed "determinantal varieties	in general, $\mathcal{M}$ not Euclidean closed
CNNs	$\mathcal{M}$ Euclidean closed, but in general, not Zariski closed	$\mathcal{M}$ Zariski closed