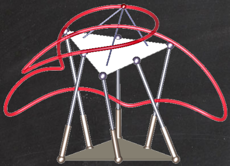
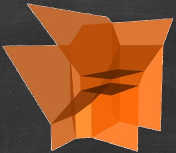
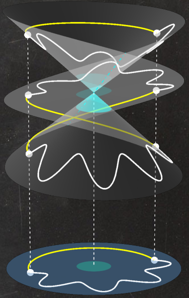


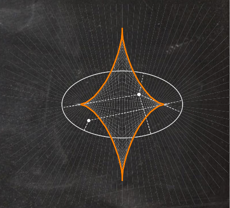




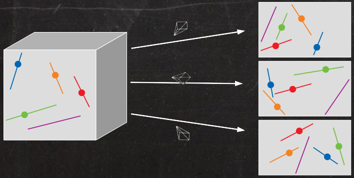
What is Nonlinear Algebra?

Kathlén Kohn

KTH Stockholm



June 10, 2020



Linear algebra

All undergraduate students learn about Gaussian elimination, a general method for solving linear systems of algebraic equations:

Input:

$$\begin{aligned} x + 2y + 3z &= 5 \\ 7x + 11y + 13z &= 17 \\ 19x + 23y + 29z &= 31 \end{aligned}$$

Output:

$$\begin{aligned} x &= -35/18 \\ y &= 2/9 \\ z &= 13/6 \end{aligned}$$

Solving very large linear systems is central to applied mathematics.

Non-linear algebra

Lucky students also learn about Gröbner bases, a general method for non-linear systems of algebraic equations:

Input:

$$\begin{aligned} x^2 + y^2 + z^2 &= 2 \\ x^3 + y^3 + z^3 &= 3 \\ x^4 + y^4 + z^4 &= 4 \end{aligned}$$

Non-linear algebra

Lucky students also learn about Gröbner bases, a general method for non-linear systems of algebraic equations:

Input:

$$x^2+y^2+z^2=2$$
$$x^3+y^3+z^3=3$$
$$x^4+y^4+z^4=4$$

Output: $3z^{12}-12z^{10}-12z^9+12z^8+72z^7-66z^6-12z^4+12z^3-1=0$

$$\begin{aligned}4y^2+(36z^{11}+54z^{10}-69z^9-252z^8-216z^7+573z^6+72z^5\\ -12z^4-99z^3+10z+3)\cdot y+36z^{11}+48z^{10}-72z^9\\ -234z^8-192z^7+564z^6-48z^5+96z^4-96z^3+10z^2+8=0\end{aligned}$$

$$\begin{aligned}4x+4y+36z^{11}+54z^{10}-69z^9-252z^8-216z^7\\ +573z^6+72z^5-12z^4-99z^3+10z+3=0\end{aligned}$$

This is very hard for large systems, but ...

The world is non-linear!

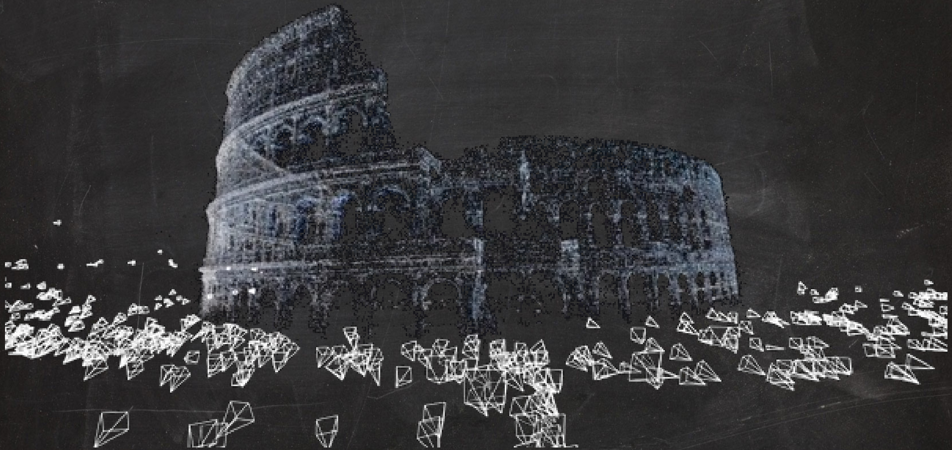
Many models in the sciences and engineering are characterized by polynomial equations. Such a set is an algebraic variety $X \subset \mathbb{R}^n$.

- computer vision
- algebraic statistics
- machine learning
- optimization
- ...

Computer Vision

Structure from Motion

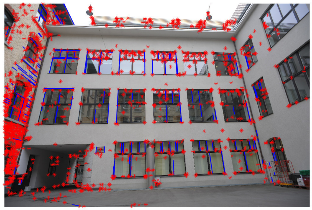
Reconstruct 3D scenes and camera poses from 2D images

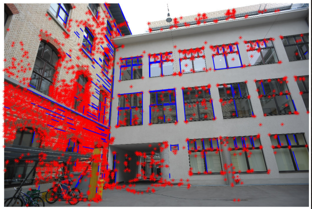


Rome in a Day: S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, S. Seitz, R. Szeliski

Reconstruct 3D scenes and camera poses from 2D images

- Step 1: Identify common points and lines on given images





- Step 2: Reconstruct coordinates of 3D points and lines as well as camera poses

Reconstruct 3D scenes and camera poses from 2D images

- Step 1: Identify common points and lines on given images

- Step 2: Reconstruct coordinates of 3D points and lines as well as camera poses

$\Rightarrow$ This is an algebraic problem!

What is a camera?

A **camera** is a 3×4 matrix C which takes pictures of points in projective 3-space via

$$\begin{aligned} \mathbb{P}^3 &\longrightarrow \mathbb{P}^2, \\ P &\longmapsto CP. \end{aligned}$$

What is a camera?

A **camera** is a 3×4 matrix C which takes pictures of points in projective 3-space via

$$\begin{aligned} \mathbb{P}^3 &\longrightarrow \mathbb{P}^2, \\ P &\longmapsto CP. \end{aligned}$$

- Each camera matrix C is a point in $\mathbb{P}^{11}$.

What is a camera?

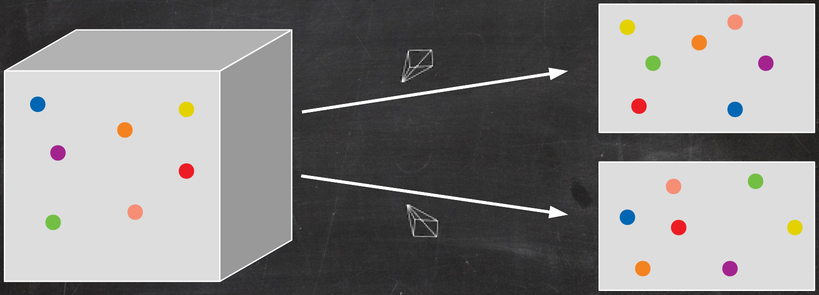
A **camera** is a 3×4 matrix C which takes pictures of points in projective 3-space via

$$\begin{aligned} \mathbb{P}^3 &\longrightarrow \mathbb{P}^2, \\ P &\longmapsto CP. \end{aligned}$$

- Each camera matrix C is a point in $\mathbb{P}^{11}$.
- There can be restrictions on the camera matrix C, e.g. by assuming that the focal length of the camera is known.

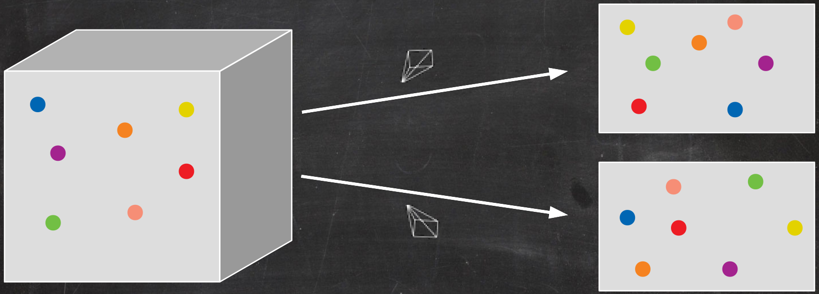
7-point problem

Given 2 images of 7 points,
can we recover the 7 points in 3D and the 2 cameras?



7-point problem

Given 2 images of 7 points,
can we recover the 7 points in 3D and the 2 cameras?

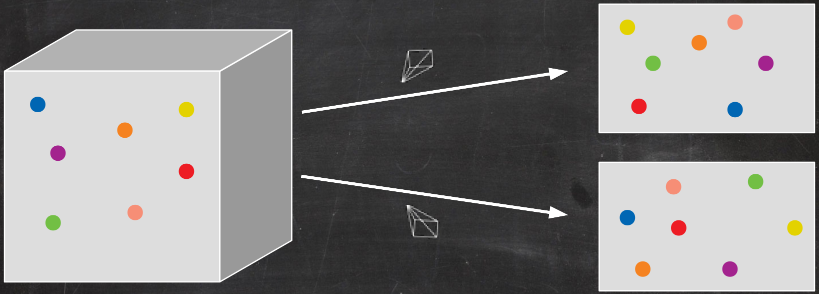


Formally, we study the **joint camera map**

$$\Phi : \left(\mathbb{P}^3\right)^7 \times \left(\mathbb{P}^{11}\right)^2 \dashrightarrow \left(\mathbb{P}^2\right)^{14},$$
$$(P_1, \ldots, P_7, C_1, C_2) \longmapsto (C_1 P_1, \ldots, C_1 P_7, C_2 P_1, \ldots, C_2 P_7),$$

7-point problem

Given 2 images of 7 points,
can we recover the 7 points in 3D and the 2 cameras?



Formally, we study the **joint camera map**

$$\Phi : \left(\mathbb{P}^3\right)^7 \times \left(\mathbb{P}^{11}\right)^2 \dashrightarrow \left(\mathbb{P}^2\right)^{14},$$
$$(P_1, \ldots, P_7, C_1, C_2) \longmapsto (C_1P_1, \ldots, C_1P_7, C_2P_1, \ldots, C_2P_7),$$

and given a point in its image $x \in \left(\mathbb{P}^2\right)^{14}$ we ask for its **fiber** $\Phi^{-1}(x)$.

7-point problem

The projective linear group $\mathrm{PGL}(4)$ acts on the fibers $\Phi^{-1}(x)$ via

$$g.(P_1,\ldots,P_7,C_1,C_2) = \left(gP_1,\ldots,gP_7,C_1g^{-1},C_2g^{-1}\right).$$

Practically, this means that we can only hope to recover points and cameras **up to projective transformations**.

joint camera map:

$$\Phi : \quad \left((\mathbb{P}^3)^7 \times (\mathbb{P}^{11})^2\right) \dashrightarrow (\mathbb{P}^2)^{14}$$

7-point problem

The projective linear group $\mathrm{PGL}(4)$ acts on the fibers $\Phi^{-1}(x)$ via

$$g.(P_1,\ldots,P_7,C_1,C_2) = \left(gP_1,\ldots,gP_7,C_1g^{-1},C_2g^{-1}\right).$$

Practically, this means that we can only hope to recover points and cameras **up to projective transformations**.

- we can adapt the joint camera map:

$$\Phi: \quad \left(\left(\mathbb{P}^3\right)^7 \times \left(\mathbb{P}^{11}\right)^2\right)/\,\mathrm{PGL}(4) \quad \dashrightarrow \quad \left(\mathbb{P}^2\right)^{14}$$

7-point problem

The projective linear group $\mathrm{PGL}(4)$ acts on the fibers $\Phi^{-1}(x)$ via

$$g.(P_1,\ldots,P_7,C_1,C_2) = \left(gP_1,\ldots,gP_7,C_1g^{-1},C_2g^{-1}\right).$$

Practically, this means that we can only hope to recover points and cameras **up to projective transformations**.

- we can adapt the joint camera map:

$$\begin{array}{rccc} \Phi: & \left((\mathbb{P}^3)^7 \times (\mathbb{P}^{11})^2\right)/\mathrm{PGL}(4) & \dashrightarrow & (\mathbb{P}^2)^{14} \\ \text{dimension:} & 3\cdot 7 + 11\cdot 2 - 15 & = 28 = & 2\cdot 14 \end{array}$$

- its fibers are **generically finite**!

7-point problem

The projective linear group $\mathrm{PGL}(4)$ acts on the fibers $\Phi^{-1}(x)$ via

$$g.(P_1, \ldots, P_7, C_1, C_2) = \left(gP_1, \ldots, gP_7, C_1 g^{-1}, C_2 g^{-1}\right).$$

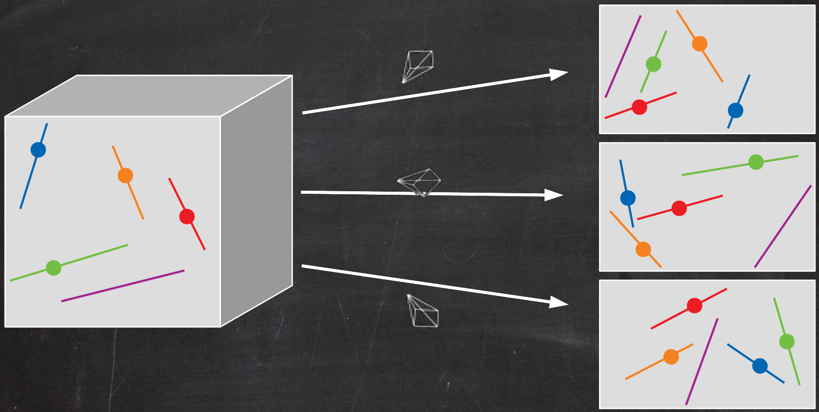
Practically, this means that we can only hope to recover points and cameras **up to projective transformations**.

- we can adapt the joint camera map:

$$\begin{array}{rccc} \Phi: & \left((\mathbb{P}^3)^7 \times (\mathbb{P}^{11})^2\right) / \mathrm{PGL}(4) & \dashrightarrow & (\mathbb{P}^2)^{14} \\ \text{dimension:} & 3 \cdot 7 + 11 \cdot 2 - 15 & = 28 = & 2 \cdot 14 \end{array}$$

- its fibers are **generically finite**!
  in fact, over $\mathbb{C}$, there are generically **3** solutions to the 7-point problem
- solving naively: **28** quadratic equations in **28** unknowns

A more complicated finite problem



Incidences are modeled by **flag varieties:** $\mathcal{F}_k := \{(P, L) \in \mathbb{P}^k \times \mathrm{Gr}(1, \mathbb{P}^k) \mid P \in L\}$

joint camera map:

$$\Phi : \left(\mathrm{Gr}(1, \mathbb{P}^3) \times (\mathcal{F}_3)^4 \times \left(\mathbb{P}^{11}\right)^3\right) / \mathrm{PGL}(4) \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^2)^3 \times (\mathcal{F}_2)^{12}$$

Algebraic Statistics

Reconstruct probability distributions from moments

Central question:
Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions on $\mathbb{R}^d$. Can we recover a distribution in the family if we know enough of its **moments**?

$$m_{i_1 i_2 \dots i_d}(\mu_\theta) = \int_{\mathbb{R}^d} w_1^{i_1} w_2^{i_2} \cdots w_d^{i_d} \, \mathrm{d}\mu_\theta \quad \text{for } i_1, i_2, \dots, i_d \in \mathbb{Z}_{\geq 0}$$

Reconstruct probability distributions from moments

Central question:

Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions on $\mathbb{R}^d$. Can we recover a distribution in the family if we know enough of its **moments**?

$$m_{i_1 i_2 \ldots i_d}(\mu_\theta) = \int_{\mathbb{R}^d} w_1^{i_1} w_2^{i_2} \cdots w_d^{i_d} \, \mathrm{d}\mu_\theta \quad \text{for } i_1, i_2, \ldots, i_d \in \mathbb{Z}_{\geq 0}$$

Example:

Let $\Theta = \{(a, b) \in \mathbb{R}^2 \mid a \leq b\}$ be the space of **line segments** in $\mathbb{R}$.
Let $\mu_{(a,b)}$ be the uniform probability distributions on the line segment (a, b).

Reconstruct probability distributions from moments

Central question:

Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions on $\mathbb{R}^d$. Can we recover a distribution in the family if we know enough of its **moments**?

$$m_{i_1 i_2 \dots i_d}(\mu_\theta) = \int_{\mathbb{R}^d} w_1^{i_1} w_2^{i_2} \cdots w_d^{i_d} \,\mathrm{d}\mu_\theta \quad \text{for } i_1, i_2, \dots, i_d \in \mathbb{Z}_{\geq 0}$$

Example:

Let $\Theta = \{(a,b) \in \mathbb{R}^2 \mid a \leq b\}$ be the space of **line segments** in $\mathbb{R}$.
Let $\mu_{(a,b)}$ be the uniform probability distributions on the line segment (a,b).

$$\Rightarrow m_i(\mu_{(a,b)}) = \frac{1}{b-a}\int_a^b w^i \,\mathrm{d}w = \frac{1}{i+1}\frac{b^{i+1} - a^{i+1}}{b-a}$$
$$= \frac{1}{i+1}\left(a^i + a^{i-1}b + a^{i-2}b^2 + \ldots + b^i\right)$$

Reconstruct probability distributions from moments

Central question:

Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions on $\mathbb{R}^d$. Can we recover a distribution in the family if we know enough of its **moments**?

$$m_{i_1 i_2 \dots i_d}(\mu_\theta) = \int_{\mathbb{R}^d} w_1^{i_1} w_2^{i_2} \cdots w_d^{i_d} \,\mathrm{d}\mu_\theta \quad \text{for } i_1, i_2, \dots, i_d \in \mathbb{Z}_{\geq 0}$$

Example:

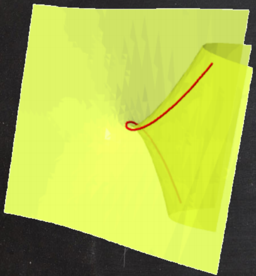
Let $\Theta = \{(a, b) \in \mathbb{R}^2 \mid a \leq b\}$ be the space of **line segments** in $\mathbb{R}$.
Let $\mu_{(a,b)}$ be the uniform probability distributions on the line segment (a, b).

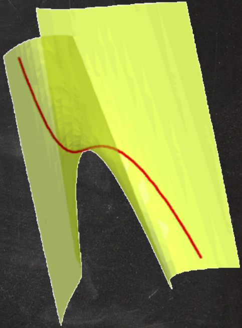
$$\Rightarrow m_i(\mu_{(a,b)}) = \frac{1}{b-a} \int_a^b w^i \,\mathrm{d}w = \frac{1}{i+1} \frac{b^{i+1} - a^{i+1}}{b-a}$$
$$= \frac{1}{i+1} \left(a^i + a^{i-1} b + a^{i-2} b^2 + \ldots + b^i\right)$$

The first two moments m_1, m_2 yield two solutions (a, b),
but only one with $a \leq b$.

Example: line segments

The moments $m_1, m_2, \ldots, m_r$ of a line segment (a, b) are not algebraically independent! **The lie on a surface in $\mathbb{R}^r$.**

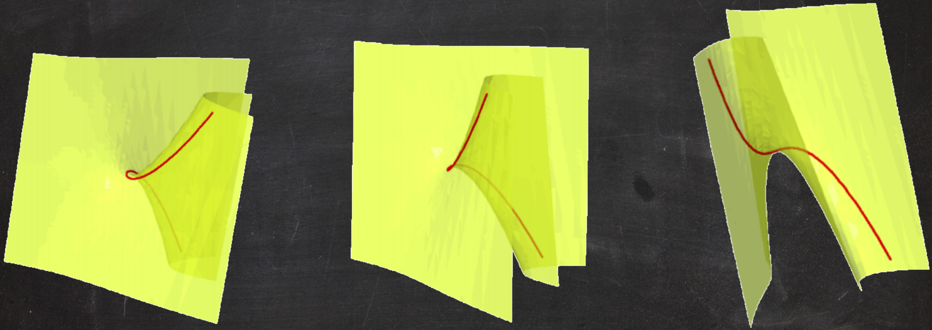




- for $r = 3$, the surface is defined by $2m_1^3 - 3m_1 m_2 + m_3 = 0$

Example: line segments

The moments $m_1, m_2, \ldots, m_r$ of a line segment (a, b) are not algebraically independent! **The lie on a surface in $\mathbb{R}^r$.**



- for $r = 3$, the surface is defined by $2m_1^3 - 3m_1m_2 + m_3 = 0$
- it contains the **twisted cubic curve** corresponding to degenerate line segments (a, a) of length 0

Example: line segments

The moments $m_1, m_2, \dots, m_r$ of a line segment (a, b) are not algebraically independent! **The lie on a surface in $\mathbb{R}^r$.**

Practical meaning:
If the given moments have noise, we cannot recover the line segment!

Example: line segments

The moments $m_1, m_2, \ldots, m_r$ of a line segment (a, b) are not algebraically independent! **The lie on a surface in $\mathbb{R}^r$.**

Practical meaning:
If the given moments have noise, we cannot recover the line segment!
We first need to denoise the moments, i.e. find a closest point on the moment surface.

Example: line segments

The moments $m_1, m_2, \ldots, m_r$ of a line segment (a, b) are not algebraically independent! **The lie on a surface in $\mathbb{R}^r$.**

Practical meaning:
If the given moments have noise, we cannot recover the line segment! We first need to denoise the moments, i.e. find a closest point on the moment surface.

$\Rightarrow$ **We need to understand the moment surface, i.e. the algebraic dependencies among the moments**.

Example: line segments

The moment surface in $\mathbb{R}^r$ of the first r moments $m_1, m_2, \ldots, m_r$

- has degree $\binom{r}{2}$
- and its prime ideal is generated by the 3×3 minors of

$$\begin{pmatrix} 0 & 1 & 2m_1 & 3m_2 & 4m_3 & \cdots & (r-1)m_{r-2} \\ 1 & 2m_1 & 3m_2 & 4m_3 & 5m_4 & \cdots & r\, m_{r-1} \\ 2m_1 & 3m_2 & 4m_3 & 5m_4 & 6m_5 & \cdots & (r+1)m_r \end{pmatrix}.$$

- These cubics form a Gröbner basis.

Intermezzo: Optimization

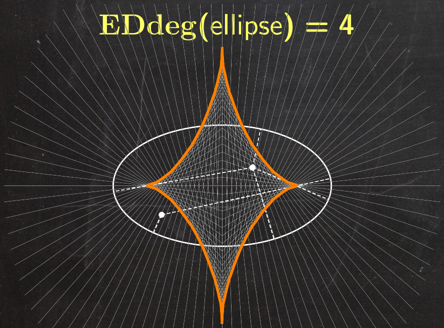
finding a closest point on an algebraic variety

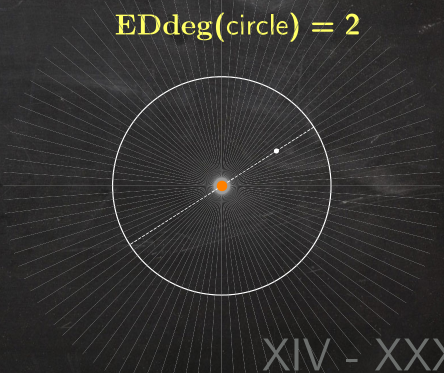
Euclidean distance degree

The **ED degree** of an algebraic variety $X \subset \mathbb{R}^n$ is the number of critical points (over $\mathbb{C}$) of the Euclidean distance

$$\begin{aligned} X &\longrightarrow \mathbb{R}, \\ x &\longmapsto \|x - u\|^2 \end{aligned}$$

between a generic point $u \in \mathbb{R}^n$ and the variety X.





back to

Algebraic Statistics

Reconstruct probability distributions from moments

Central question:

Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions on $\mathbb{R}^d$. Can we recover a distribution in the family if we know enough of its **moments**?

Similarities to reconstruction in computer vision:

Instead of the joint camera map, we study the **moment map**

$$\begin{aligned} \Phi : \Theta &\longrightarrow \mathbb{R}^{\mathcal{I}}, \\ \theta &\longmapsto m_{i_1 i_2 \ldots i_d}(\mu_\theta), \end{aligned}$$

where $\mathcal{I} \subset \mathbb{Z}_{\geq 0}$ is a finite index set, and ask for its fibers.

Reconstruct probability distributions from moments

Central question:

Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions on $\mathbb{R}^d$. Can we recover a distribution in the family if we know enough of its **moments**?

Similarities to reconstruction in computer vision:

Instead of the joint camera map, we study the **moment map**

$$\begin{aligned} \Phi : \Theta &\longrightarrow \mathbb{R}^{\mathcal{I}}, \\ \theta &\longmapsto m_{i_1 i_2 \dots i_d}(\mu_\theta), \end{aligned}$$

where $\mathcal{I} \subset \mathbb{Z}_{\geq 0}$ is a finite index set, and ask for its fibers.

Typical settings in practice:

1. the fibers of Φ are generically finite and non-empty
   $\rightarrow$ can solve reconstruction problem for any generic input
2. $\mathrm{im}(\Phi)$ lies in a proper subvariety
   $\rightarrow$ need to denoise input before reconstructing

Example: quadrilaterals

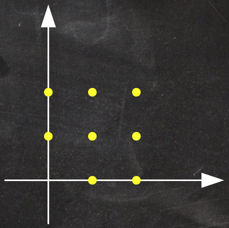
Let $\Theta = \{\Box \subset \mathbb{R}^2\} \subset \left(\mathbb{R}^2\right)^4$ be the space of **quadrilaterals** in $\mathbb{R}^2$.
Let $\mu_\Box$ be the uniform probability distribution on the quadrilateral $\Box$.

Example: quadrilaterals

Let $\Theta = \{\square \subset \mathbb{R}^2\} \subset \left(\mathbb{R}^2\right)^4$ be the space of **quadrilaterals** in $\mathbb{R}^2$.
Let $\mu_\square$ be the uniform probability distribution on the quadrilateral $\square$.

Let $\mathcal{I}$ be as shown on the right.

The fibers of $\Phi : \Theta \to \mathbb{R}^8$ are generically finite, of cardinality 80 over $\mathbb{C}$.



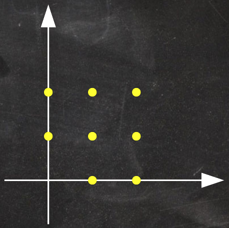
Example: quadrilaterals

Let $\Theta = \{\square \subset \mathbb{R}^2\} \subset \left(\mathbb{R}^2\right)^4$ be the space of **quadrilaterals** in $\mathbb{R}^2$.
Let $\mu_\square$ be the uniform probability distribution on the quadrilateral $\square$.

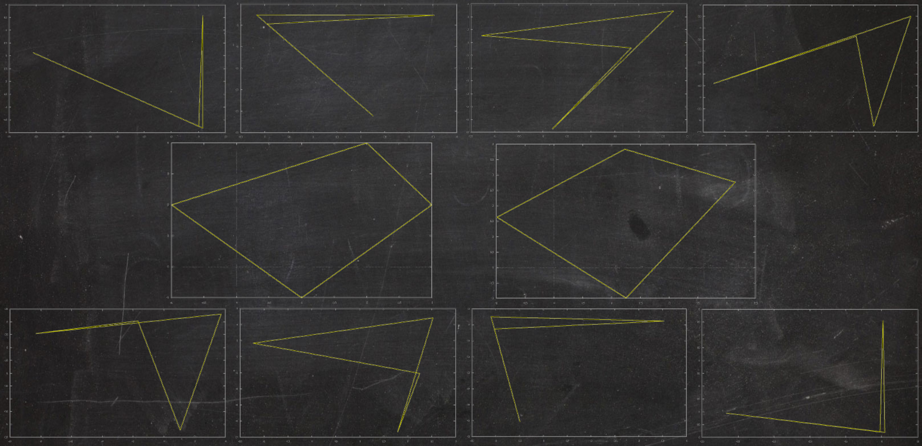
Let $\mathcal{I}$ be as shown on the right.

The fibers of $\Phi : \Theta \to \mathbb{R}^8$ are generically finite, of cardinality 80 over $\mathbb{C}$.

The dihedral group of order 8 acts on each fiber.
$\rightsquigarrow$ Each fiber consists of 10 "quadrilaterals".



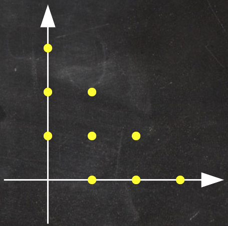
Example: quadrilaterals



Example: quadrilaterals

Let $\Theta = \{\square \subset \mathbb{R}^2\} \subset \left(\mathbb{R}^2\right)^4$ be the space of **quadrilaterals** in $\mathbb{R}^2$.
Let $\mu_\square$ be the uniform probability distribution on the quadrilateral $\square$.

Let $\mathcal{I}$ be as shown on the right.



The Zariski closure of the image of $\Phi : \Theta \to \mathbb{R}^9$ is a hypersurface.

We can compute it using the **invariant ring** of the **affine group** $\mathrm{Aff}(\mathbb{R}^2)$.

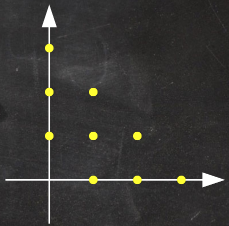
Example: quadrilaterals

Let $\Theta = \{\square \subset \mathbb{R}^2\} \subset \left(\mathbb{R}^2\right)^4$ be the space of **quadrilaterals** in $\mathbb{R}^2$.
Let $\mu_\square$ be the uniform probability distribution on the quadrilateral $\square$.

Let $\mathcal{I}$ be as shown on the right.

The Zariski closure of the image of $\Phi : \Theta \to \mathbb{R}^9$ is a hypersurface.

We can compute it using the **invariant ring** of the **affine group** $\mathrm{Aff}(\mathbb{R}^2)$.



The moment hypersurface has degree 18.
Its defining polynomial has 5100 terms.

Example: quadrilaterals

```
         6   12                  3   2   12              2   4   12                  5   11                    4       11                      2   2   11              2       3   11
2125764m02 m10   - 1574640m00*m02 m03 m10   + 291600m00 m03 m10   - 25509168m01*m02 m10  m11 + 9447840m00*m02 m03*m10  m11 + 9447840m00*m01*m02 m03 m10  m11 - 3499200m00 m02*m03 m10  m11 +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
           2   4   10   2                5   10   2                      3       10   2                2       2   10   2              2   2   2   10   2              2       3   10   2
127545840m01 m02 m10  m11  + 8030664m00*m02 m10  m11  - 75582720m00*m01*m02 m03*m10  m11  - 18895680m00*m01 m02*m03 m10  m11  + 7523280m00 m02 m03 m10  m11  + 6998400m00 m01*m03 m10  m11  -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
           3   3   9   3                   4   9   3                     2   2       9   3                2   3       9   3              3   2   9   3              2           2   9   3
340122240m01 m02 m10 m11  - 80306640m00*m01*m02 m10 m11  + 226748160m00*m01 m02 m03*m10 m11  + 30443040m00 m02 m03*m10 m11  + 12597120m00*m01 m03 m10 m11  - 30093120m00 m01*m02*m03 m10 m11  -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
         3   3   9   3              4   2   8   4              2   3   8   4              2   4   8   4                  3           8   4              2       2       8   4
4665600m00 m03 m10 m11  + 510183360m01 m02 m10 m11  + 321226560m00*m01 m02 m10 m11  + 8503056m00 m02 m10 m11  - 302330880m00*m01 m02*m03*m10 m11  - 182658240m00 m01*m02 m03*m10 m11  +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
          2   2   2   8   4              3       2   8   4              5       7   5                  3   2   7   5              2       3   7   5                  4       7   5
30093120m00 m01 m03 m10 m11  + 27643680m00 m02*m03 m10 m11  - 408146688m01 m02*m10 m11  - 642453120m00*m01 m02 m10 m11  - 68024448m00 m01*m02 m10 m11  + 151165440m00*m01 m03*m10 m11  +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
           2   2           7   5              3   2       7   5              3       2   7   5              6   6   6                  4       6   6              2   2   2   6   6
365316480m00 m01 m02*m03*m10 m11  + 25894080m00 m02 m03*m10 m11  - 55287360m00 m01*m03 m10 m11  + 136048896m01 m10 m11  + 642453120m00*m01 m02*m10 m11  + 204073344m00 m01 m02 m10 m11  +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
        3   3   6   6              2   3       6   6              3               6   6              4   2   6   6                  5   5   7              2   3       5   7
186624m00 m02 m10 m11  - 243544320m00 m01 m03*m10 m11  - 103576320m00 m01*m02*m03*m10 m11  + 19215360m00 m03 m10 m11  - 256981248m00*m01 m10 m11  - 272097792m00 m01 m02*m10 m11  -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
         3       2   5   7              3   2       5   7              4           5   7              2   4   4   8              3   2       4   8              4   2   4   8
1119744m00 m01*m02 m10 m11  + 103576320m00 m01 m03*m10 m11  - 518400m00 m02*m03*m10 m11  + 136048896m00 m01 m10 m11  + 2239488m00 m01 m02*m10 m11  - 2985984m00 m02 m10 m11  +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
         4           4   8              3   3   3   9              4           3   9              5       3   9              4   2   2   10              5       2   10              5           11
1036800m00 m01*m03*m10 m11  - 1492992m00 m01 m10 m11  + 11943936m00 m01*m02*m10 m11  - 4423680m00 m03*m10 m11  - 11943936m00 m01 m10 m11   - 663552m00 m02*m10 m11   + 1327104m00 m01*m10*m11   -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
            5   11                       3       11                  2   2   2   11                2       3   11                      4   10                        2   2       10
9447840m00*m02 m10  m12 + 9447840m00*m01*m02 m03*m10  m12 + 3499200m00 m02 m03 m10  m12 - 3499200m00 m01*m03 m10  m12 + 66134880m00*m01*m02 m10  m11*m12 - 56687040m00*m01 m02 m03*m10  m11*m12 -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
          2   3       10                   2           2   10                  3   3   10                      2   3   9   2              2   4   9   2
30443040m00 m02 m03*m10  m11*m12 + 17496000m00 m01*m02*m03 m10  m11*m12 + 3499200m00 m03 m10  m11*m12 - 151165440m00*m01 m02 m10 m11 m12 - 22517352m00 m02 m10 m11 m12 +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
              3           9   2                  2       2       9   2              2   2   2   9   2              3       2   9   2                  3   2   8   3
113374080m00*m01 m02*m03*m10 m11 m12 + 137518560m00 m01*m02 m03*m10 m11 m12 - 48988800m00 m01 m03 m10 m11 m12 - 19245600m00 m02*m03 m10 m11 m12 + 75582720m00*m01 m02 m10 m11 m12 +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
          2       3   8   3                  4       8   3                  2   2           8   3              3   2       8   3              3       2   8   3
88809696m00 m01*m02 m10 m11 m12 - 75582720m00*m01 m03*m10 m11 m12 - 184757760m00 m01 m02*m03*m10 m11 m12 - 56337120m00 m02 m03*m10 m11 m12 + 80481600m00 m01*m03 m10 m11 m12 +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
              4       7   4                2   2   2   7   4              3   3   7   4              2   3       7   4              3               7   4              4   2   7   4
151165440m00*m01 m02*m10 m11 m12 + 7558272m00 m01 m02 m10 m11 m12 - 5388768m00 m02 m10 m11 m12 + 62985600m00 m01 m03*m10 m11 m12 + 59486400m00 m01*m02*m03*m10 m11 m12 - 30002400m00 m03 m10 m11 m12 -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
              5   6   5                2   3       6   5              3       2   6   5              3   2       6   5              4           6   5              2   4   5   6
151165440m00*m01 m10 m11 m12 - 375394176m00 m01 m02*m10 m11 m12 - 45349632m00 m01*m02 m10 m11 m12 + 106375680m00 m01 m03*m10 m11 m12 - 4043520m00 m02*m03*m10 m11 m12 + 370355328m00 m01 m10 m11 m12 +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
           3   2       5   6              4   2   5   6              4           5   6              3   3   4   7              4           4   7              5       4   7
246063744m00 m01 m02*m10 m11 m12 + 13281408m00 m02 m10 m11 m12 - 107205120m00 m01*m03*m10 m11 m12 - 267618816m00 m01 m10 m11 m12 - 51570432m00 m01*m02*m10 m11 m12 + 19388160m00 m03*m10 m11 m12 +
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
          4   2   3   8              5       3   8              5       2   9              6       10                  2   3   10   2              2   4   10   2
50015232m00 m01 m10 m11 m12 + 3981312m00 m02*m10 m11 m12 + 5308416m00 m01*m10 m11 m12 - 663552m00 m10*m11  m12 - 14171760m00*m01 m02 m10  m12  + 17557236m00 m02 m10  m12  -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
          2       2       10   2              2   2   2   10   2              3       2   10   2              3   2   9       2              2       3   9       2
20995200m00 m01*m02 m03*m10  m12  + 15746400m00 m01 m03 m10  m12  - 2624400m00 m02*m03 m10  m12  + 85030560m00*m01 m02 m10 m11*m12  - 49128768m00 m01*m02 m10 m11*m12  -
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
          2   2           9       2              3   2       9       2              3       2   9       2              4       8   2   2              2   2   2   8   2   2
```

Find distribution best explaining data

Central question:

- Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions.
- Let $Y = (Y_1, \ldots, Y_n)$ be n samples of observed data.

Can we find a distribution in the family that best fits the empirical data Y?

Find distribution best explaining data

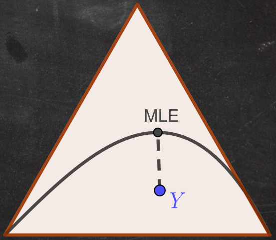
Central question:

- Let $\{\mu_\theta \mid \theta \in \Theta\}$ be a family of probability distributions.
- Let $Y = (Y_1, \ldots, Y_n)$ be n samples of observed data.

Can we find a distribution in the family that best fits the empirical data Y?

Approach: maximize the **likelihood function**

$$L_Y(\theta) := \mu_\theta(Y_1) \cdots \mu_\theta(Y_n), \qquad \text{where } \theta \in \Theta.$$



A **maximum likelihood estimate (MLE)** is a distribution in the family that maximizes the likelihood L_Y.

Example: (conditional) independence

Consider two random variables X and Y having m and n states.
Their joint probability distribution is an $m \times n$ matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}$$

whose entries are non-negative and sum to 1.

Example: (conditional) independence

Consider two random variables X and Y having m and n states.

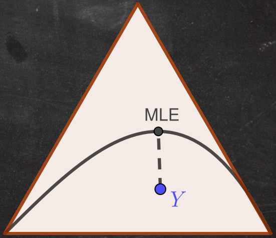
Their joint probability distribution is an $m \times n$ matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}$$

whose entries are non-negative and sum to 1.

Let $\mathcal{M}_r$ be the variety of rank-r matrices in the probability simplex Δ_{mn-1}.

Matrices P in $\mathcal{M}_1$ represent **independent distributions**.



Example: (conditional) independence

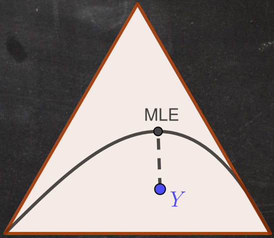
Consider two random variables X and Y having m and n states.
Their joint probability distribution is an $m \times n$ matrix

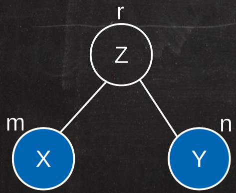
$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}$$

whose entries are non-negative and sum to 1.

Let $\mathcal{M}_r$ be the variety of rank-r matrices in the probability simplex Δ_{mn-1}.

Matrices P in $\mathcal{M}_1$ represent **independent distributions**.





$\mathcal{M}_r$ comprises mixtures of r independent distributions.
Its elements P represent **conditionally independent distributions**.

XXI - XXX

Example: (conditional) independence

Suppose i.i.d. samples are drawn from an unknown distribution.
We summarize these data also in a matrix

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}.$$

Example: (conditional) independence

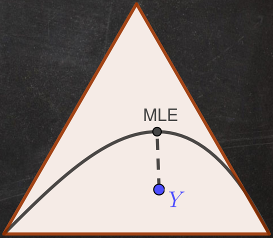
Suppose i.i.d. samples are drawn from an unknown distribution.
We summarize these data also in a matrix

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}.$$

The likelihood function is the monomial

$$L_Y(P) = \prod_{i=1}^{m} \prod_{j=1}^{n} p_{ij}^{y_{ij}}.$$

An MLE for data Y is a rank-r matrix $P \in \mathcal{M}_r$ maximizing $L_Y(P)$.



ML degree

The **ML degree** of a family of distributions is the number of critical points (over $\mathbb{C}$) of the likelihood function for generic data.

some known[1] ML degrees of the rank varieties $\mathcal{M}_r$:

$(m,n) =$	$(3,3)$	$(3,4)$	$(3,5)$	$(4,4)$	$(4,5)$	$(4,6)$	$(5,5)$
$r=1$	1	1	1	1	1	1	1
$r=2$	**10**	**26**	**58**	**191**	**843**	**3119**	**6776**
$r=3$	1	1	1	**191**	**843**	**3119**	**61326**
$r=4$				1	1	1	**6776**
$r=5$							1

[1] Hauenstein, Rodriguez, Sturmfels: *Maximum likelihood for matrices with rank constraints*, Journal of Algebraic Statistics **5** (2014) 18–38.

Machine Learning

Neural networks

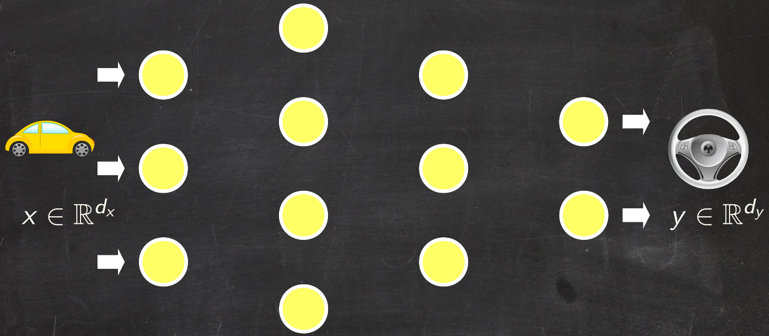
XXIV - XXX

Neural networks

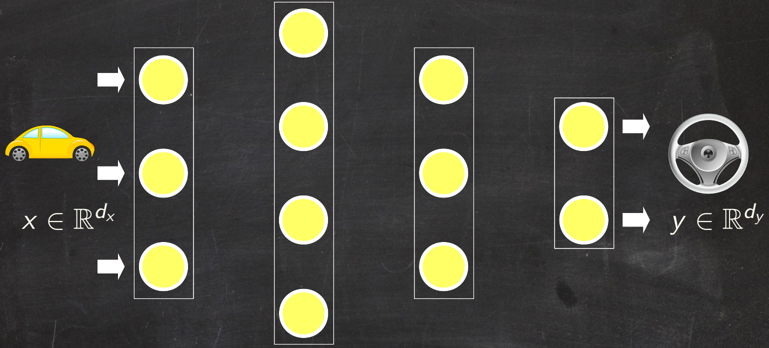
$x \in \mathbb{R}^{d_x}$

$y \in \mathbb{R}^{d_y}$

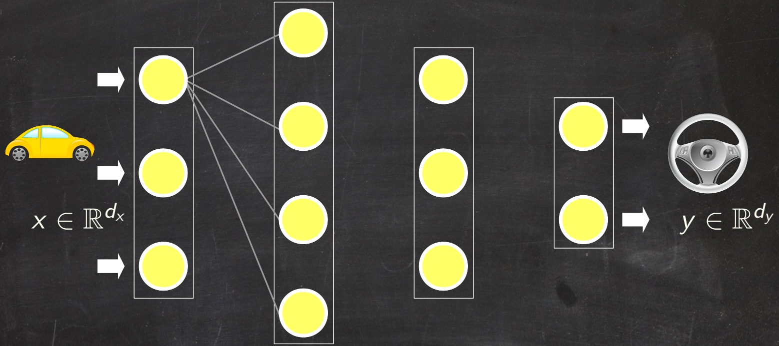
Neural networks



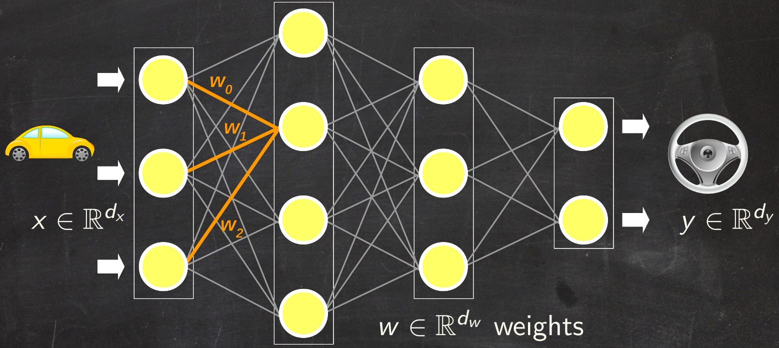
Neural networks



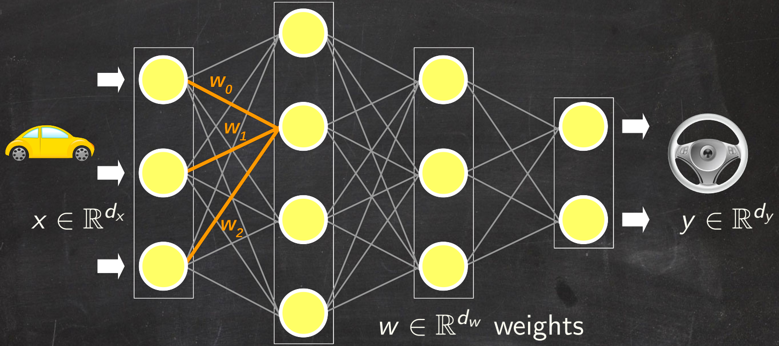
Neural networks



Neural networks

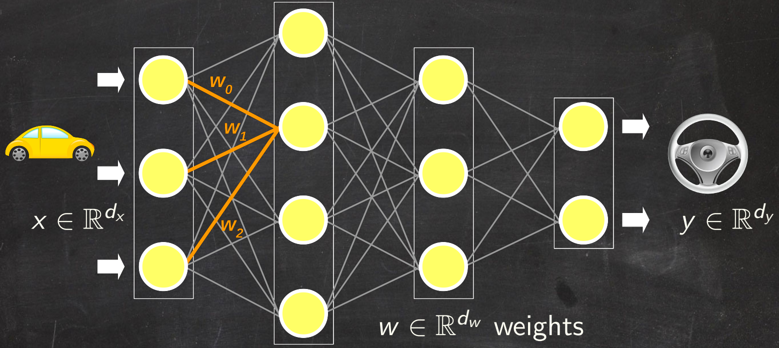


Neural networks



A neural network is defined by a continuous mapping $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$.

Neural networks

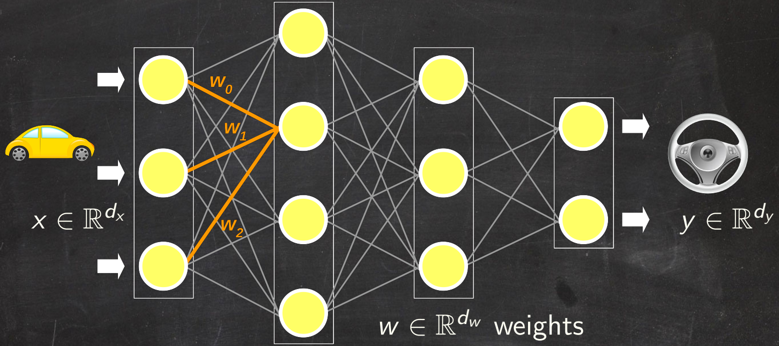


A neural network is defined by a continuous mapping $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$.

Definition $\mathcal{M}_\Phi := \left\{ \Phi(w, \cdot) : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y} \mid w \in \mathbb{R}^{d_w} \right\}$

is called the **neuromanifold** of Φ.

Neural networks



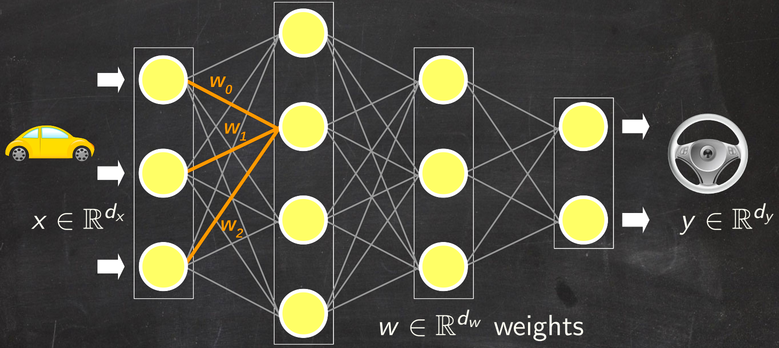
A neural network is defined by a continuous mapping $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$.

Definition $\mathcal{M}_\Phi := \left\{ \Phi(w, \cdot) : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y} \mid w \in \mathbb{R}^{d_w} \right\}$

is called the **neuromanifold** of Φ.

Observation 1. Φ piecewise smooth $\Rightarrow \mathcal{M}_\Phi$ manifold with singularities

Neural networks



A neural network is defined by a continuous mapping $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$.

Definition $\mathcal{M}_\Phi := \left\{ \Phi(w, \cdot) : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y} \mid w \in \mathbb{R}^{d_w} \right\} \subset C(\mathbb{R}^{d_x}, \mathbb{R}^{d_y})$

is called the **neuromanifold** of Φ.

Observation
1. Φ piecewise smooth $\Rightarrow \mathcal{M}_\Phi$ manifold with singularities
2. $\dim \mathcal{M}_\Phi \leq d_w$

Linear networks

A **linear network** is defined by a map $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$ of the form

$$\Phi(w, x) = W_h W_{h-1} \dots W_1 x,$$
$$\text{where } w = (W_h, \dots, W_1) \text{ and } W_i \in \mathbb{R}^{d_i \times d_{i-1}},$$

(so $d_w = d_h d_{h-1} + \ldots + d_1 d_0$, $d_x = d_0$ and $d_y = d_h$).

Linear networks

A **linear network** is defined by a map $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$ of the form

$$\Phi(w, x) = W_h W_{h-1} \dots W_1 x,$$
$$\text{where } w = (W_h, \dots, W_1) \text{ and } W_i \in \mathbb{R}^{d_i \times d_{i-1}},$$

(so $d_w = d_h d_{h-1} + \dots + d_1 d_0$, $d_x = d_0$ and $d_y = d_h$).

Example
The neuromanifold of the linear network Φ is the bounded rank variety

$$\mathcal{M}_\Phi = \left\{ M \in \mathbb{R}^{d_h \times d_0} \mid \mathrm{rk}(M) \leq \underbrace{\min\{d_0, d_1, \dots, d_h\}}_{=:r} \right\}.$$

Loss landscapes

A **loss function** on a neural network $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$ is of the form

$$\begin{aligned} L : \mathbb{R}^{d_w} &\xrightarrow{\quad \mu \quad} \mathcal{M}_\Phi \xrightarrow{\quad \ell|_{\mathcal{M}_\Phi} \quad} \mathbb{R}, \\ w &\longmapsto \Phi(w, \cdot) \end{aligned}$$

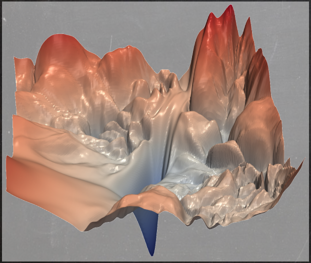
where ℓ is a functional defined on a subset of $C(\mathbb{R}^{d_x}, \mathbb{R}^{d_y})$ containing $\mathcal{M}_\Phi$.

Loss landscapes

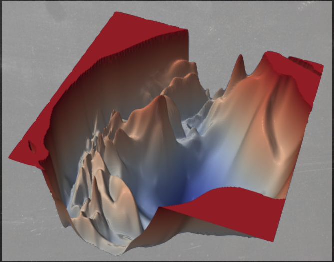
A **loss function** on a neural network $\Phi : \mathbb{R}^{d_w} \times \mathbb{R}^{d_x} \longrightarrow \mathbb{R}^{d_y}$ is of the form

$$\begin{array}{rcccl} L : \mathbb{R}^{d_w} & \xrightarrow{\quad\mu\quad} & \mathcal{M}_\Phi & \xrightarrow{\quad\ell|_{\mathcal{M}_\Phi}\quad} & \mathbb{R}, \\ w & \longmapsto & \Phi(w,\cdot) & & \end{array}$$

where ℓ is a functional defined on a subset of $C(\mathbb{R}^{d_x}, \mathbb{R}^{d_y})$ containing $\mathcal{M}_\Phi$.



Visualizations of L



Source: Li, Hao, et al. "Visualizing the loss landscape of neural nets." Advances in Neural Information Processing Systems. 2018.

Quadratic loss on linear networks

Fixed data matrices $X \in \mathbb{R}^{d_0 \times s}$ and $Y \in \mathbb{R}^{d_h \times s}$ define a **quadratic loss**

$$\begin{aligned} \ell_{X,Y} : \mathbb{R}^{d_h \times d_0} &\longrightarrow \mathbb{R}, \\ M &\longmapsto \|MX - Y\|_F^2 \end{aligned}$$

Quadratic loss on linear networks

Fixed data matrices $X \in \mathbb{R}^{d_0 \times s}$ and $Y \in \mathbb{R}^{d_h \times s}$ define a **quadratic loss**

$$\begin{aligned} \ell_{X,Y} : \mathbb{R}^{d_h \times d_0} &\longrightarrow \mathbb{R}, \\ M &\longmapsto \|MX - Y\|_F^2 \end{aligned}$$

Observation If $XX^T = \mathrm{I}_{d_0}$ ("whitened data"), then

$$\ell_{X,Y}(M) = \|M - YX^T\|_F^2 + \text{const.}$$

Quadratic loss on linear networks

Fixed data matrices $X \in \mathbb{R}^{d_0 \times s}$ and $Y \in \mathbb{R}^{d_h \times s}$ define a **quadratic loss**

$$\begin{aligned} \ell_{X,Y} : \mathbb{R}^{d_h \times d_0} &\longrightarrow \mathbb{R}, \\ M &\longmapsto \|MX - Y\|_F^2 \end{aligned}$$

Observation If $XX^T = \mathrm{I}_{d_0}$ ("whitened data"), then

$$\ell_{X,Y}(M) = \|M - YX^T\|_F^2 + \mathrm{const.}$$

Minimizing $\ell_{X,Y}$ on the bounded rank variety $\mathcal{M}_\Phi = \{M \mid \mathrm{rk}(M) \leq r\}$ is equivalent to minimizing the Euclidean distance of YX^T to $\mathcal{M}_\Phi$.

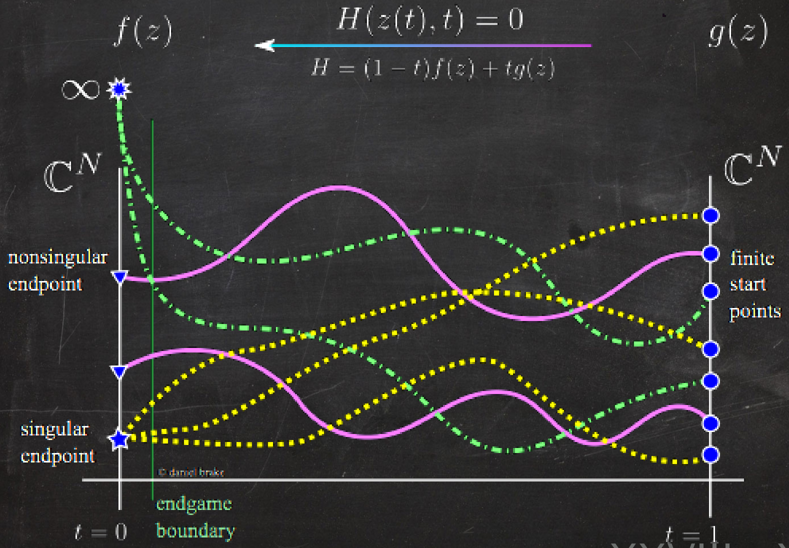
How to

solve systems of polynomial equations?

(besides Gröbner bases)

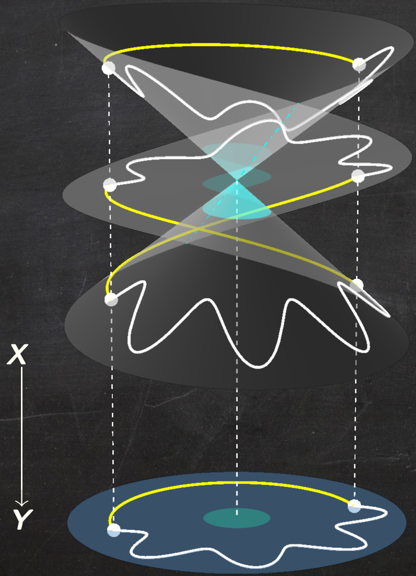
Numerical algebraic geometry

homotopy continuation



Numerical algebraic geometry

monodromy



Numerical algebraic geometry

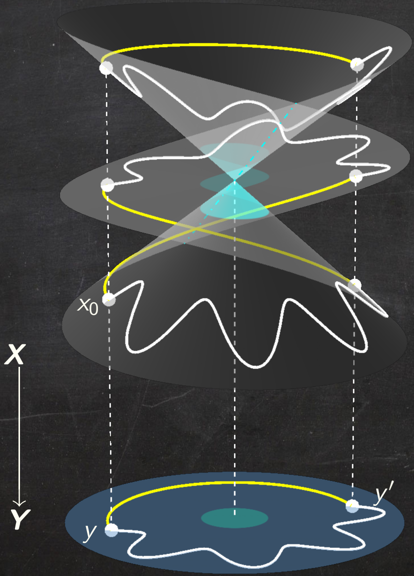
monodromy

Numerical algebraic geometry

monodromy

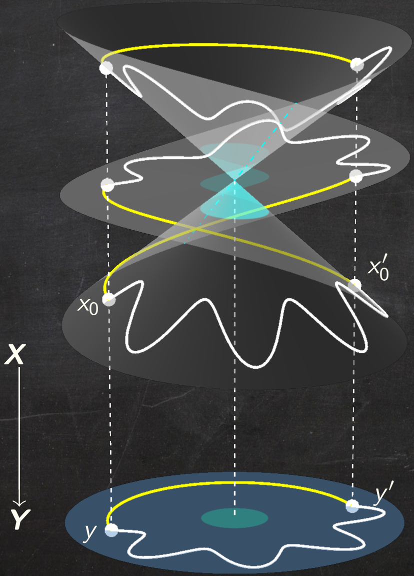
Numerical algebraic geometry

monodromy



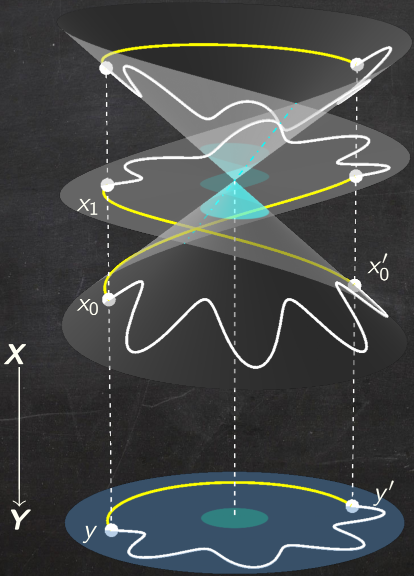
Numerical algebraic geometry

monodromy



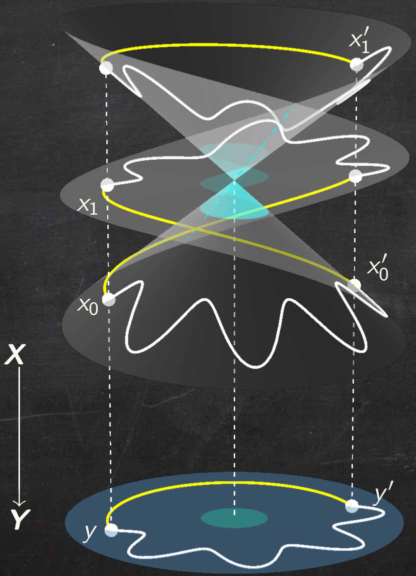
Numerical algebraic geometry

monodromy



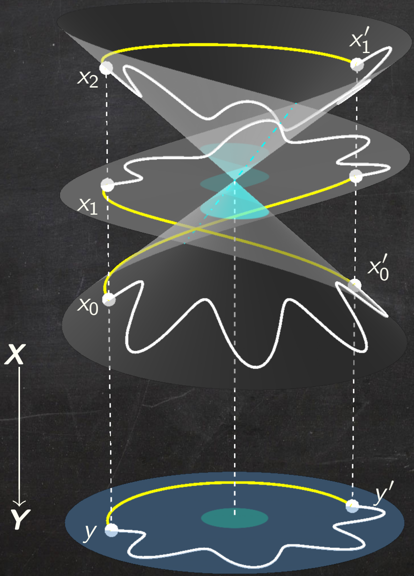
Numerical algebraic geometry

monodromy



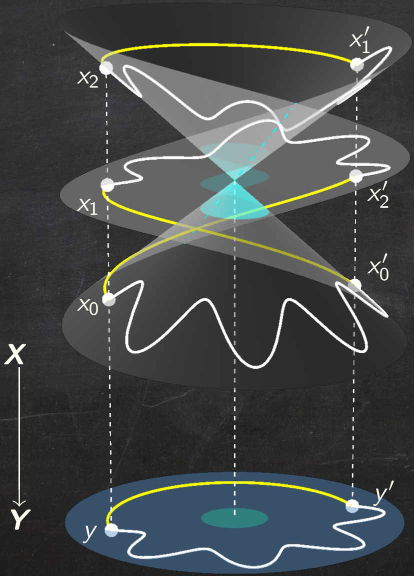
Numerical algebraic geometry

monodromy



Numerical algebraic geometry

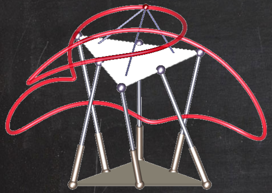
monodromy



The world is non-linear!

Application areas

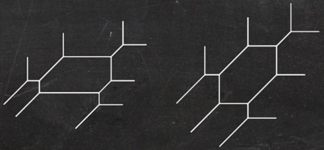
- computer vision
- algebraic statistics
- machine learning
- optimization
- robotics



- complexity theory
- biochemistry
- music
- ...

Toolbox

- algebraic geometry
- combinatorics
- convex and discrete geometry
- representation theory
- symbolic and numerical computations
- tropical geometry



- ...

XXX - XXX