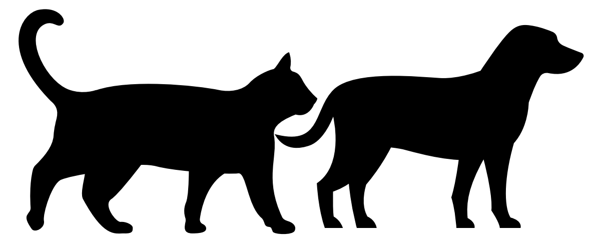
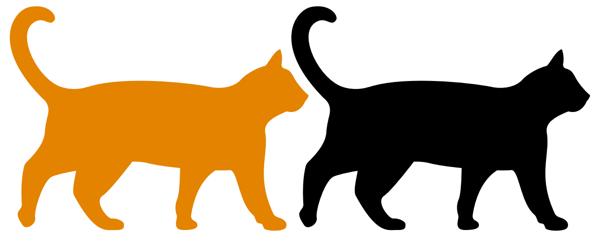
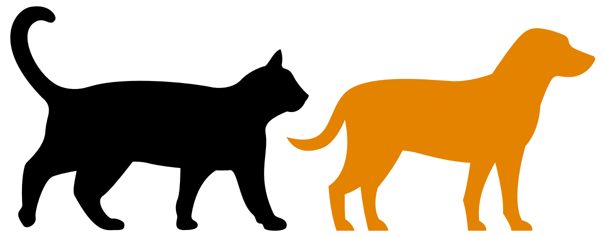
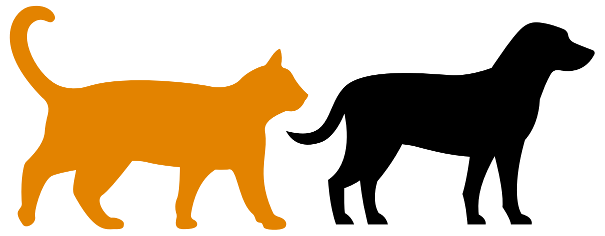
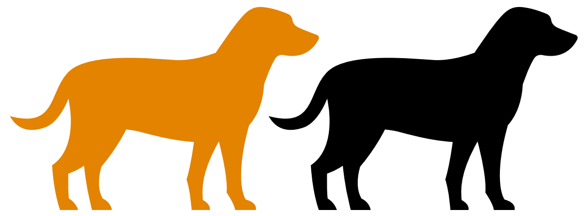
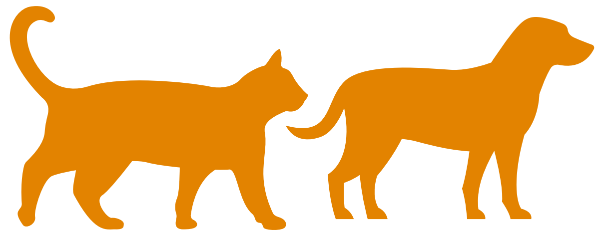
Likelihood Geometry of the Squared Grassmannian

Hannah Friedman (UC Berkeley)
Algebraic Statistics in our Changing World at the Joint Math Meetings 2025
January 8, 2025

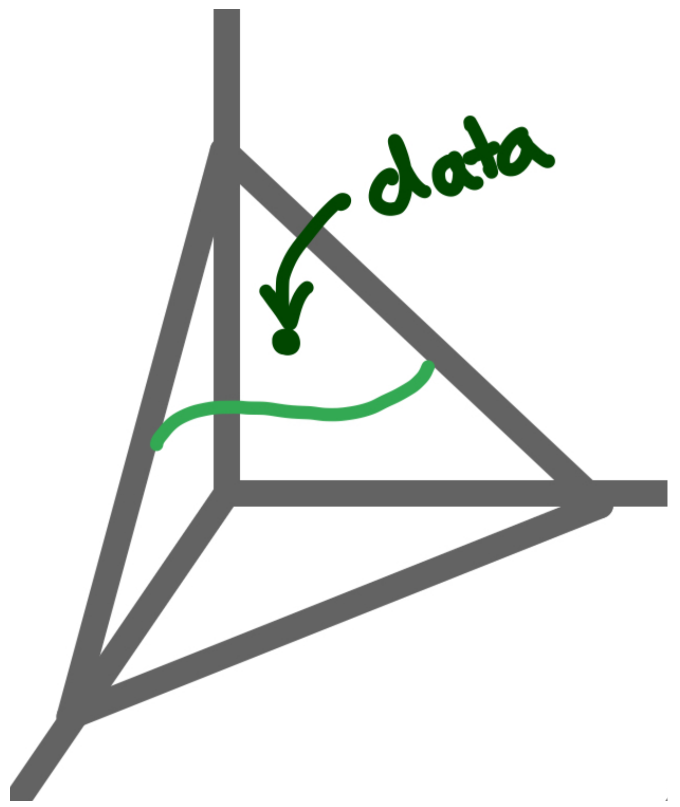
Jackie walks into an animal shelter and adopts 2 of the 4 animals at the shelter every day for 100 days. Every day, she decides which animals to take home by sampling from an unknown probability distribution.

Maximum Likelihood Estimation

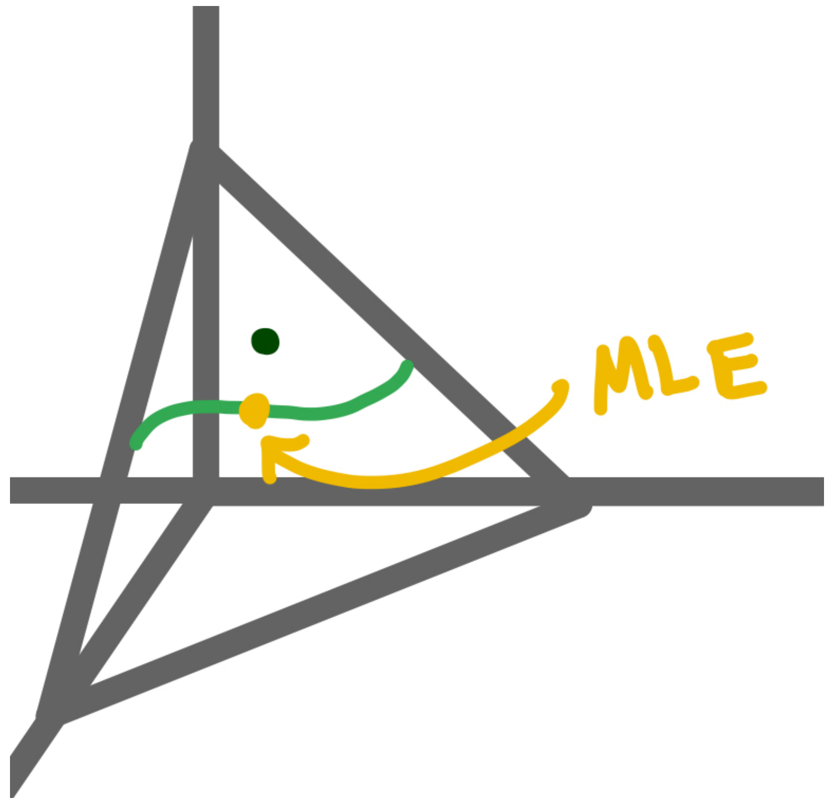
	14
	11
	26
	24
	9
	16

Since Jackie prefers to take home a pair of different animals, we assume that Jackie is sampling from a specific type of distribution called a **projection determinantal point process** (projection DPP).

Given:



Find:



Projection Determinantal Point Processes

Example Projection DPPs with state space $\binom{[4]}{2}$ are parameterized by symmetric matrices

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{12} & p_{22} & p_{23} & p_{24} \\ p_{13} & p_{23} & p_{33} & p_{34} \\ p_{14} & p_{24} & p_{34} & p_{44} \end{pmatrix}$$

satisfying

$$P^2 = P$$
$$\mathrm{trace}(P) = 2$$

and the distribution is defined by

$$\mathbb{P}_{ij} = \det(P_{ij}) = p_{ii}p_{jj} - p_{ij}^2 .$$

The definition is the same for projection DPPs with state space $\binom{[n]}{2}$

—just take P to be an $n \times n$ matrix.

Two Lives of the Grassmannian

Definition The Grassmannian $\mathrm{Gr}(2,n)$ is the variety of 2-dimensional subspaces of $\mathbb{R}^n$.

Every point in $\mathrm{Gr}(2,n)$ is the row span of some $A \in \mathbb{R}^{2\times n}$, but this representation is not unique.

Orthogonal Projection Matrices

$$P = A^T(AA^T)^{-1}A$$

Plücker Coordinates

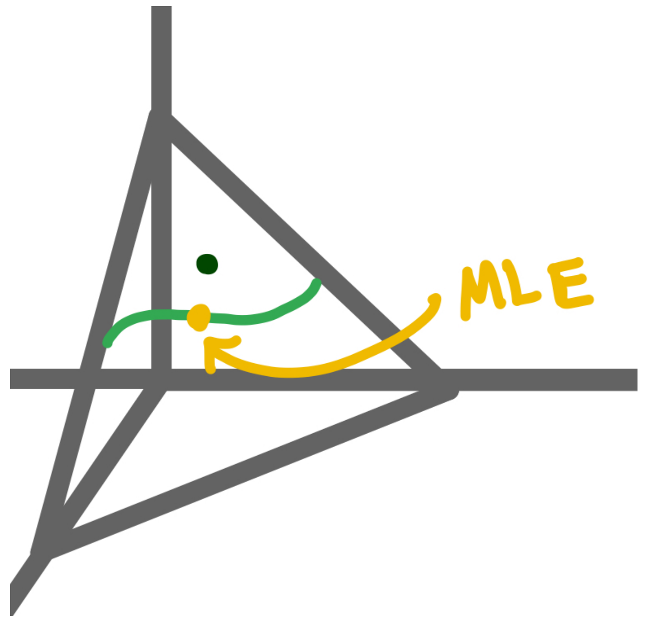
$$x = (\det(A_{ij}))_{1\le i<j\le n}$$

Lemma (Devriendt-F-Reinke-Sturmfels, 2024)

$$\mathbb{P}_{ij} = \det(P_{ij}) = \frac{x_{ij}^2}{\sum_{1\le k<\ell\le n} x_{k\ell}^2}.$$

Computing the Maximum Likelihood Estimate

Every 2-dimensional subspace of $\mathbb{R}^n$ determines a projection DPP by



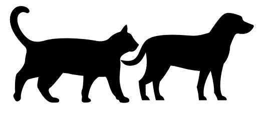
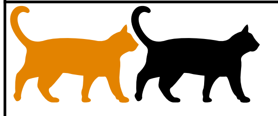
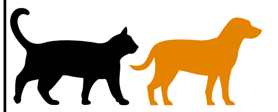
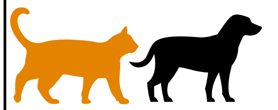
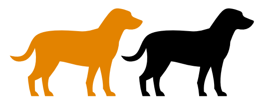
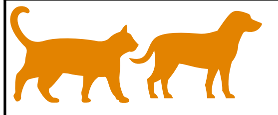
$$A = \begin{pmatrix} 1 & 0 & a_{13} & \cdots & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2n} \end{pmatrix} \qquad \mathbb{P}_{ij} = \det(P_{ij}) = \frac{\det(A_{ij})^2}{\sum_{1 \le k < \ell \le n} (A_{k\ell})^2}.$$

To compute the maximum likelihood estimate, we find the matrix A which maximizes the **log-likelihood function**

$$L_u(A) = \sum_{i,j} u_{ij} \log(\det(A_{ij})^2) - \left(\sum_{i,j} u_{ij}\right) \log\left(\sum_{i,j} \det(A_{ij})^2\right).$$

Example (n = 4)

$$A = \begin{pmatrix} 1 & 0 & a_{13} & a_{14} \\ 0 & 1 & a_{23} & a_{24} \end{pmatrix} \qquad u = [14, 11, 26, 24, 9, 16]$$

	14
	11
	26
	24
	9
	16

$$L_u(A) = 14\log(1) + 11\log(a_{23}^2) + 26\log(a_{24}^2) + 24\log(a_{13}^2) + 9\log(a_{14}^2) + 16\log((a_{13}a_{24} - a_{14}a_{23})^2)$$
$$- 100\log(1 + a_{23}^2 + a_{24}^2 + a_{13}^2 + a_{14}^2 + (a_{13}a_{24} - a_{14}a_{23})^2)$$

Computing the Maximum Likelihood Estimate

$$L_u(A) = 14\log(1) + 11\log(a_{23}^2) + 26\log(a_{24}^2) + 24\log(a_{13}^2) + 9\log(a_{14}^2) + 16\log((a_{13}a_{24} - a_{14}a_{23})^2)$$
$$-100\log(1 + a_{23}^2 + a_{24}^2 + a_{13}^2 + a_{14}^2 + (a_{13}a_{24} - a_{14}a_{23})^2)$$

1.

$$\frac{\partial L_u}{\partial a_{13}} = \frac{48}{a_{13}} + \frac{32a_{24}}{a_{13}a_{24} - a_{14}a_{23}} - 200\frac{a_{13} + a_{24}(a_{13}a_{12} - a_{14}a_{23})}{1 + a_{23}^2 + a_{24}^2 + a_{13}^2 + a_{14}^2 + (a_{13}a_{24} - a_{14}a_{23})^2} = 0$$

$$\frac{\partial L_u}{\partial a_{14}} = \frac{18}{a_{14}} - \frac{32a_{23}}{a_{13}a_{24} - a_{14}a_{23}} - 200\frac{a_{14} - a_{23}(a_{13}a_{12} - a_{14}a_{23})}{1 + a_{23}^2 + a_{24}^2 + a_{13}^2 + a_{14}^2 + (a_{13}a_{24} - a_{14}a_{23})^2} = 0$$

$$\frac{\partial L_u}{\partial a_{23}} = \frac{22}{a_{23}} - \frac{32a_{14}}{a_{13}a_{24} - a_{14}a_{23}} - 200\frac{a_{23} - a_{14}(a_{13}a_{12} - a_{14}a_{23})}{1 + a_{23}^2 + a_{24}^2 + a_{13}^2 + a_{14}^2 + (a_{13}a_{24} - a_{14}a_{23})^2} = 0$$

$$\frac{\partial L_u}{\partial a_{24}} = \frac{52}{a_{24}} + \frac{32a_{13}}{a_{13}a_{24} - a_{14}a_{23}} - 200\frac{a_{24} + a_{13}(a_{13}a_{12} - a_{14}a_{23})}{1 + a_{23}^2 + a_{24}^2 + a_{13}^2 + a_{14}^2 + (a_{13}a_{24} - a_{14}a_{23})^2} = 0$$

2. Apply **monodromy_solve** in **HomotopyContinuation.jl**.

$$\begin{pmatrix}1 & 0 & 1.308 & 0.802\\ 0 & 1 & 0.886 & 1.361\end{pmatrix} \begin{pmatrix}1 & 0 & 1.308 & -0.802\\ 0 & 1 & -0.886 & 1.361\end{pmatrix} \begin{pmatrix}1 & 0 & -1.308 & -0.802\\ 0 & 1 & 0.886 & 1.361\end{pmatrix} \begin{pmatrix}1 & 0 & 1.308 & -0.802\\ 0 & 1 & 0.886 & -1.361\end{pmatrix}$$

$$\begin{pmatrix}1 & 0 & -1.308 & -0.802\\ 0 & 1 & -0.886 & -1.361\end{pmatrix} \begin{pmatrix}1 & 0 & -1.308 & 0.802\\ 0 & 1 & 0.886 & -1.361\end{pmatrix} \begin{pmatrix}1 & 0 & 1.308 & 0.802\\ 0 & 1 & -0.886 & -1.361\end{pmatrix} \begin{pmatrix}1 & 0 & -1.308 & 0.802\\ 0 & 1 & -0.886 & 1.361\end{pmatrix}$$

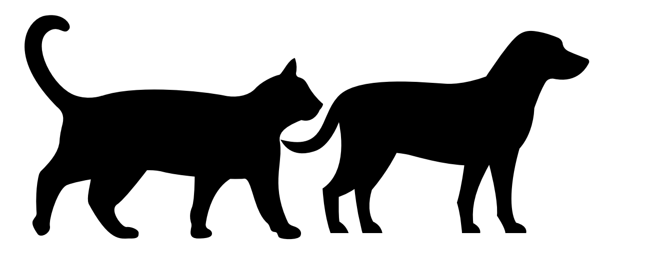
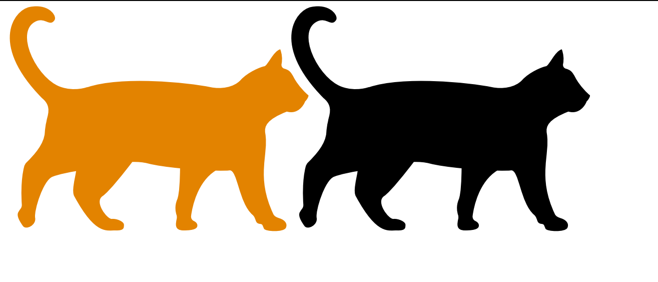
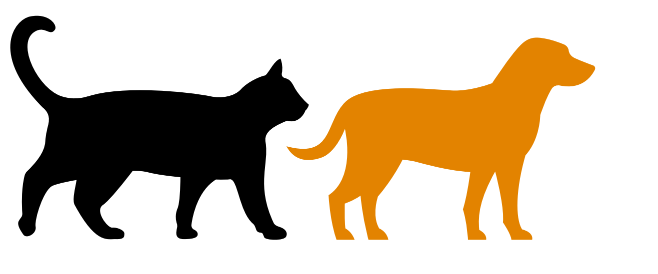
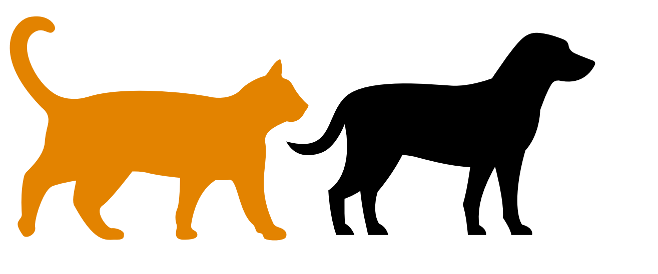
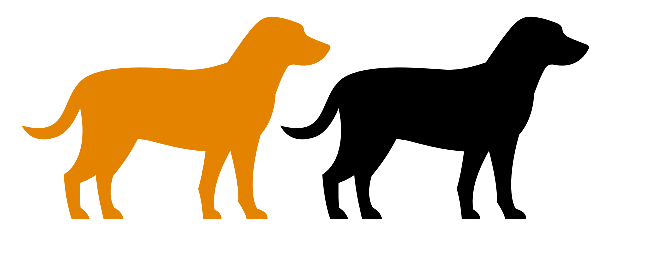
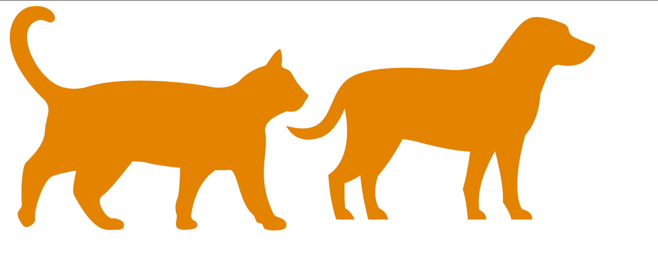
$$\begin{pmatrix}1 & 0 & 0.839 & -0.507\\ 0 & 1 & 0.584 & 0.888\end{pmatrix} \times 8 \qquad \begin{pmatrix}1 & 0 & 1.320 & 1.690\\ 0 & 1 & 1.759 & 1.408\end{pmatrix} \times 8$$

24 critical points

$\det(A_{ij})^2$ →

$$\begin{pmatrix}1\\0.786\\1.852\\1.710\\0.643\\1.143\end{pmatrix}, \begin{pmatrix}1\\0.341\\0.788\\0.704\\0.257\\1.083\end{pmatrix}, \begin{pmatrix}1\\3.093\\1.982\\1.744\\2.855\\1.238\end{pmatrix}$$

Three Kinds of MLEs

	14
	11
	26
	24
	9
	16

$$A^* = \begin{pmatrix} 1 & 0 & 1.308 & 0.802 \\ 0 & 1 & 0.886 & 1.361 \end{pmatrix}$$

(unique up to flipping some signs)

$$P^* = \begin{pmatrix} 0.51 & -0.3154 & 0.3872 & -0.0204 \\ -0.3154 & 0.47 & 0.0041 & 0.3867 \\ 0.3872 & 0.0041 & 0.51 & 0.3161 \\ -0.0204 & 0.3867 & 0.3161 & 0.51 \end{pmatrix}$$

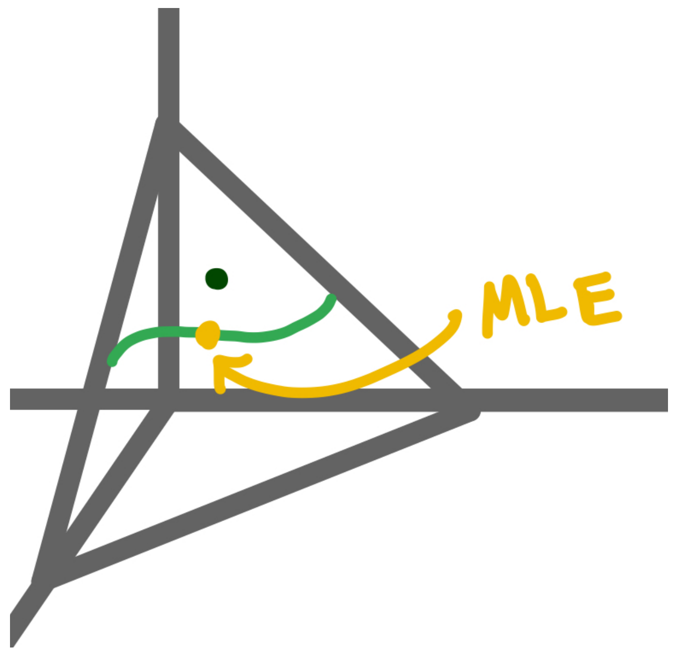
(unique up to flipping some signs)

$$q^* = \begin{pmatrix} 1 \\ 0.786 \\ 1.852 \\ 1.710 \\ 0.643 \\ 1.143 \end{pmatrix} \sim \begin{pmatrix} 0.14 \\ 0.110 \\ 0.259 \\ 0.239 \\ 0.090 \\ 0.160 \end{pmatrix}$$

(unique)

The Squared Grassmannian…

…is a model for projection DPPs!



Definition The **squared Grassmannian** $\mathrm{sGr}(2,n)$ is the image of the Grassmannian $\mathrm{Gr}(2,n) \subset \mathbb{P}^{\binom{n}{2}-1}$ in its Plücker embedding under the map

$$\begin{aligned} \mathrm{Gr}(2,n) &\to \mathbb{P}^{\binom{n}{2}-1} \\ (x_{ij})_{1\le i<j\le n} &\mapsto (x_{ij}^2)_{1\le i<j\le n} \end{aligned}$$

Corollary (Devriendt-F-Reinke-Sturmfels, 2024) The projection determinantal point process is the discrete statistical model on the state space $\binom{[n]}{2}$ whose underlying algebraic variety is the squared Grassmannian $\mathrm{sGr}(2,n)$.

$$L_u(A) = \sum_{i,j} u_{ij}\log(\det(A_{ij})^2) - \left(\sum_{i,j} u_{ij}\right)\log\left(\sum_{i,j}\det(A_{ij})^2\right) \quad \text{vs.} \quad \underset{q\,\in\,\mathrm{sGr}(2,n)}{L_u(q)} = \sum_{i,j} u_{ij}\log(q_{ij}) - \left(\sum_{ij} u_{ij}\right)\log\left(\sum_{i,j} q_{ij}\right)$$

Theorem (Huh-Sturmfels, 2014) The number of critical points of $L_u(q)$ is generically finite and does not depend on u. This number is called the **maximum likelihood degree** (ML degree) of $\mathrm{sGr}(2,n)$.

Motivating the Maximum Likelihood Degree

Example The ML degree of $\mathrm{sGr}(2,n)$ is 3: $$\begin{pmatrix} 1 \\ 0.786 \\ 1.852 \\ 1.710 \\ 0.643 \\ 1.143 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.341 \\ 0.788 \\ 0.704 \\ 0.257 \\ 1.083 \end{pmatrix}, \begin{pmatrix} 1 \\ 3.093 \\ 1.982 \\ 1.744 \\ 2.855 \\ 1.238 \end{pmatrix}$$

1. The more critical points there are, the harder the problem is to solve. The ML degree is an algebraic measure of the **difficulty of the problem**.

2. When numerically computing the solution to such an optimization problem, a heuristic stopping criterion is applied. Knowing the number of solutions a priori means that we don't need to wait until the criterion is met, so the **computation is much faster**.

Likelihood Geometry of the Squared Grassmannian

Theorem (F, 2024). The number of complex critical points of the parametric log-likelihood function

$$L_u(A) = \sum_{i,j} u_{ij} \log(\det(A_{ij})^2) - \left(\sum_{i,j} u_{ij}\right) \log \left(\sum_{i,j} \det(A_{ij})^2\right)$$

is $2^{n-2}(n-1)!$

Corollary (F, 2024). The ML degree of the squared Grassmannian $\mathrm{sGr}(2,n)$ is $\dfrac{(n-1)!}{2}$.

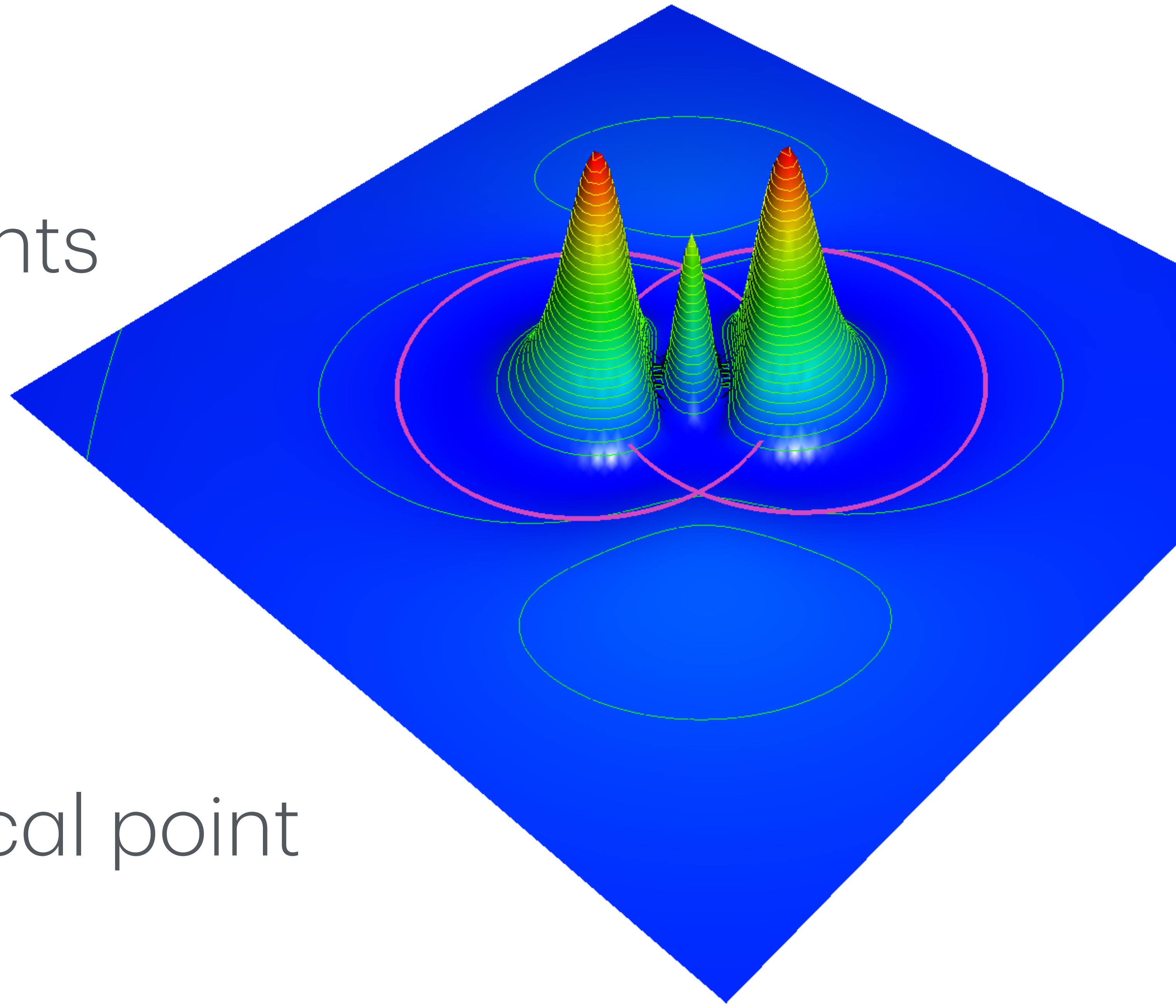
proof idea: Apply the following theorem

Theorem (Huh, 2013). If the very affine variety $X \backslash \mathscr{H}$ is smooth of dimension d, then the ML degree of X is the signed Euler characteristic $(-1)^d \chi(X \backslash \mathscr{H})$.

and compute the Euler characteristic inductively using the deletion map

$$\begin{pmatrix} 1 & 0 & a_{13} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2(n-2)} & a_{2n} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & a_{13} & \cdots & a_{1(n-1)} \\ 0 & 1 & a_{23} & \cdots & a_{2(n-2)} \end{pmatrix}$$

Real and Positive Solutions



Example Parametric Critical Points

$$\begin{pmatrix} 1 & 0 & 1.308 & 0.802 \\ 0 & 1 & 0.886 & 1.361 \end{pmatrix} \times 8$$
$$\begin{pmatrix} 1 & 0 & 0.839 & -0.507 \\ 0 & 1 & 0.584 & 0.888 \end{pmatrix} \times 8$$
$$\begin{pmatrix} 1 & 0 & 1.320 & 1.690 \\ 0 & 1 & 1.759 & 1.408 \end{pmatrix} \times 8$$

Implicit Critical points

$$\begin{pmatrix} 1 \\ 0.786 \\ 1.852 \\ 1.710 \\ 0.643 \\ 1.143 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.341 \\ 0.788 \\ 0.704 \\ 0.257 \\ 1.083 \end{pmatrix}, \begin{pmatrix} 1 \\ 3.093 \\ 1.982 \\ 1.744 \\ 2.855 \\ 1.238 \end{pmatrix}$$

Theorem (F, 2024) All critical points are real and positive. Every critical point is a local maximum of the likelihood function.

proof: Squaring means real parametric critical points imply positive critical points.

The likelihood function $\ell_u(A) = \dfrac{\prod_{i,j} \det(A_{ij})^{2u_{ij}}}{\left(\sum_{i,j} \det(A_{ij})^2\right)^{\sum_{ij} u_{i,j}}}$ is nonnegative and therefore

has at least one local maximum in every region, bounded or unbounded, of $\mathbb{R}^{2(n-2)} \setminus \bigcup_{i,j} \{\det(A_{ij}) = 0\}$.

Real and Positive Solutions

Claim. The space $\mathbb{R}^{2(n-2)} \setminus \bigcup_{i,j} \{\det(A_{ij}) = 0\}$ has $2^{n-2}(n-1)!$ connected regions.

The regions are in bijection with the possible sign vectors that can arise from a vector of Plücker coordinates in $\mathrm{Gr}(2,n)$.

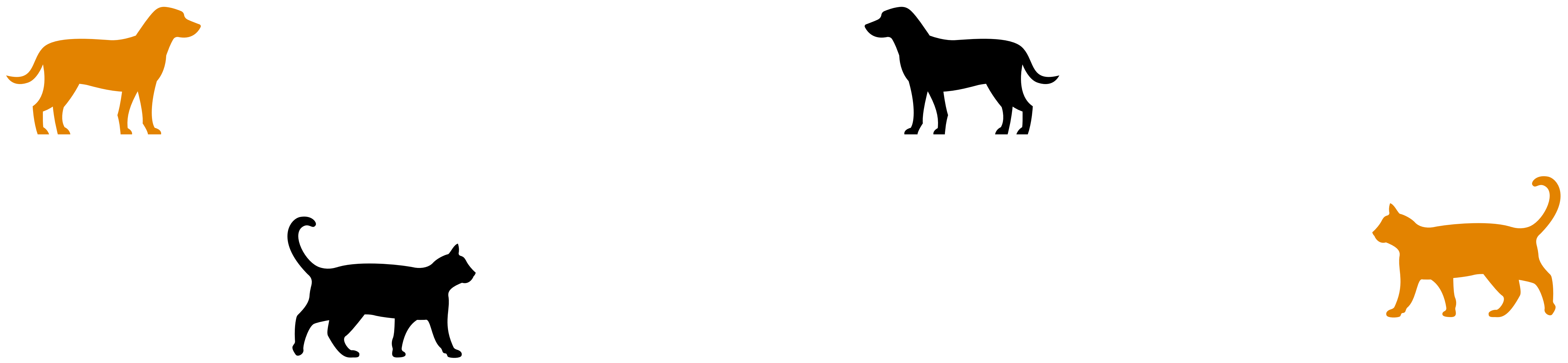
1. Choose how many columns have two different signs ($n-1$ choices).

$$A_n = \begin{pmatrix} 1 & 0 & -a_{13} & \cdots & -a_{1k} & a_{1(k+1)} & \cdots & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2k} & a_{2(k+1)} & \cdots & a_{2n} \end{pmatrix}$$

2. Permute the last $n-2$ columns ($(n-2)!$ choices).

3. Flip the signs of any of the last $n-2$ columns (2^{n-2} choices). ■

Thank you!



- Hannah Friedman, *Likelihood Geometry of the Squared Grassmannian* (2024), `arXiv: 2409.03730`.
- Karel Devriendt, Hannah Friedman, Bernhard Reinke, and Bernd Sturmfels, *The Two Lives of the Grassmannian*, to appear in *Acta Universitatis Sapientiae, Mathematica* (2024).
- June Huh, *The Maximum Likelihood Degree of a Very Affine Variety*, Composito Mathematica **149** (2013), 1245-1266.
- June Huh and Bernd Sturmfels, *Likelihood Geometry*, Combinatorial Algebraic Geometry (eds. Aldo Conca et al.), Lecture Notes in Mathematics 2108, Springer, (2014) 63-117.
- Paul Breiding and Sascha Timme, *HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia*, Mathematical Software - ICMS 2018, Spring International Publishing (2018), 458-465.