ML estimation in Gaussian graphical models

Caroline Uhler

Department of Statistics

UC Berkeley

The Second CREST-SBM International Conference, Osaka

June 28, 2010

Outline

- Background and setup:

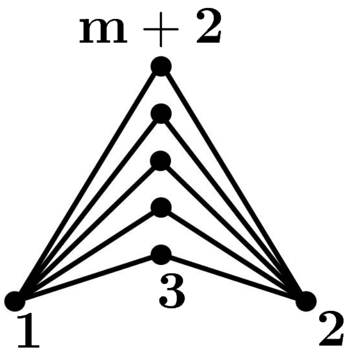
  - Gaussian graphical models

  - Maximum likelihood estimation

- Existence of the maximum likelihood estimate (MLE)
  - "Cone" problem
  - "Probability" problem

- ML degree of a graph

★ Example: $K_{2,m}$



Gaussian graphical models

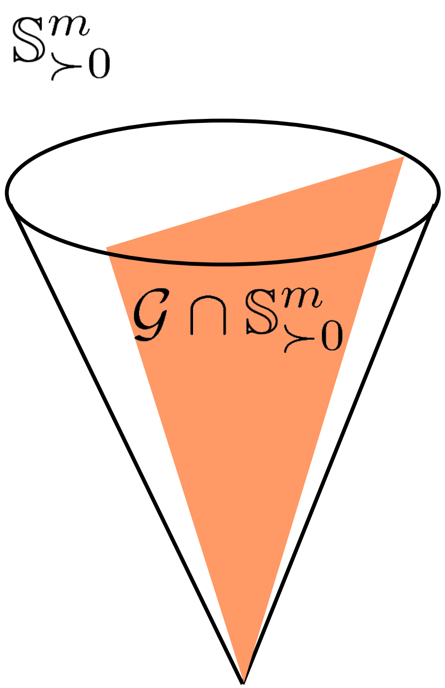
- $\mathcal{N}_{\mathbf{m}}(\mathbf{0}, \mathbf{\Sigma})$:
  - $G = ([m], E)$ undirected graph with $(\alpha, \alpha) \in E \quad \forall \alpha \in [m]$.
  - $\Sigma \in \mathbb{S}^m_{\succ 0}$ covariance matrix on $[m]$
  - $K := \Sigma^{-1} \in \mathbb{S}^m_{\succ 0}$ concentration matrix with $K \in \mathcal{G}$,

    $\mathcal{G} := \{M \in \mathbb{S}^m \ : \ M_{\alpha\beta} = 0, \quad \forall (\alpha, \beta) \notin E\}$

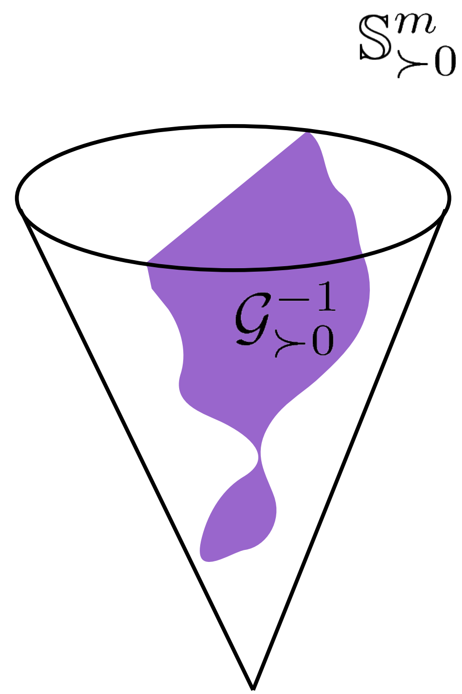
- **Gaussian graphical model:** $\mathcal{G}^{-1}_{\succ 0} := \left\{\Sigma \in \mathbb{S}^m_{\succ 0} \ : \ \Sigma^{-1} \in \mathcal{G}\right\}$

Geometry

Concentration matrices: $K = \Sigma^{-1}$



Covariance matrices: Σ



Gaussian graphical models

- $\mathcal{N}_{\mathbf{m}}(\mathbf{0}, \boldsymbol{\Sigma})$:
  - $G = ([m], E)$ undirected graph with $(\alpha, \alpha) \in E \quad \forall \alpha \in [m]$.
  - $\Sigma \in \mathbb{S}^m_{\succ 0}$ covariance matrix on $[m]$
  - $K := \Sigma^{-1} \in \mathbb{S}^m_{\succ 0}$ concentration matrix with $K \in \mathcal{G}$,

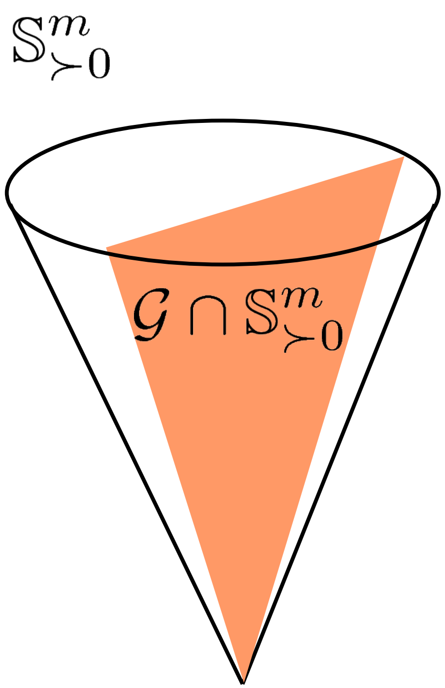
    $\mathcal{G} := \{M \in \mathbb{S}^m \ : \ M_{\alpha\beta} = 0, \quad \forall (\alpha, \beta) \notin E\}$

- **Gaussian graphical model:** $\mathcal{G}^{-1}_{\succ 0} := \left\{\Sigma \in \mathbb{S}^m_{\succ 0} \ : \ \Sigma^{-1} \in \mathcal{G}\right\}$

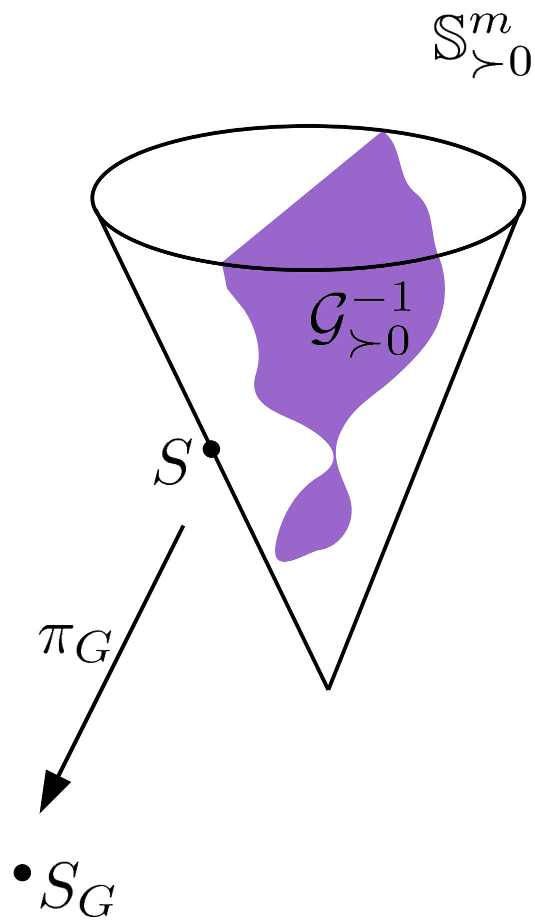
- **Data:**
  - $X_1, \ldots, X_n \in \mathbb{R}^m$ i.i.d samples from $\mathcal{N}_m(0, \Sigma), \quad n < m$
  - $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T \in \mathbb{S}^m_{\succeq 0}$ sample covariance matrix
  - $S_G := (S_{\alpha\beta})_{(\alpha,\beta) \in E}$ sufficient statistics

Geometry

Concentration matrices: $K = \Sigma^{-1}$

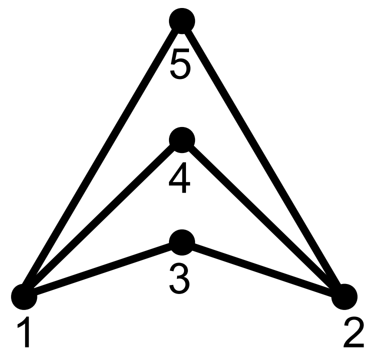


Covariance matrices: Σ



Example $K_{2,3}$

$G =$



- The corresponding Gaussian graphical model consists of multivariate Gaussians with concentration matrix of the form

$$K = \begin{pmatrix} \lambda_{11} & 0 & \lambda_{13} & \lambda_{14} & \lambda_{15} \\ 0 & \lambda_{22} & \lambda_{23} & \lambda_{24} & \lambda_{25} \\ \lambda_{13} & \lambda_{23} & \lambda_{33} & 0 & 0 \\ \lambda_{14} & \lambda_{24} & 0 & \lambda_{44} & 0 \\ \lambda_{15} & \lambda_{25} & 0 & 0 & \lambda_{55} \end{pmatrix}.$$

- Given a sample covariance matrix S, the sufficient statistics are

$$S_G = (S_{11}, S_{13}, S_{14}, S_{15}, S_{22}, S_{23}, S_{24}, S_{25}, S_{33}, S_{44}, S_{55}).$$

Maximum likelihood estimation

- Log-likelihood function:

$$\log\det(K) - \langle S, K\rangle \;=\; \log\det\Big(\sum\nolimits_{(\alpha,\beta)\in E} \lambda_{\alpha\beta}\mathbb{1}_{\alpha\beta}\Big) - \sum\nolimits_{(\alpha,\beta)\in E} \lambda_{\alpha\beta} S_{\alpha\beta}$$

Theorem *(regular exponential families)*:

In a Gaussian graphical model the MLEs $\hat{\Sigma}$ and $\hat{K}$ exist if and only if

$$\mathrm{fiber}_G(S) := \{\Sigma \in \mathbb{S}^m_{\succ 0} \mid \Sigma_{\alpha\beta} \;=\; S_{\alpha\beta},\ \forall(\alpha,\beta) \in E\} \neq \emptyset,$$

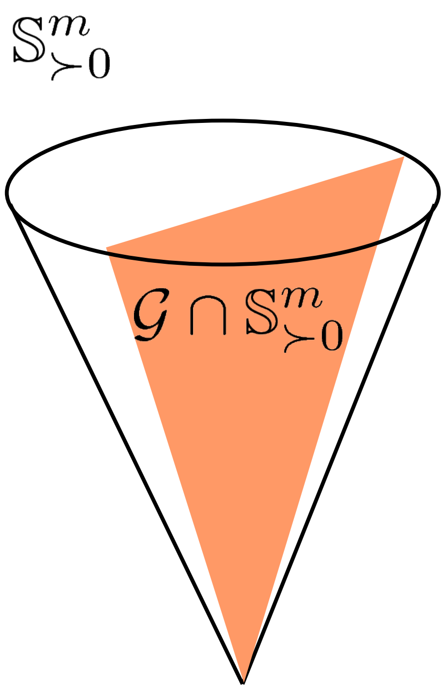
i.e. S_G is PD-completable.

Then $\hat{\Sigma} \in \mathcal{G}^{-1}_{\succ 0}$ is uniquely determined by

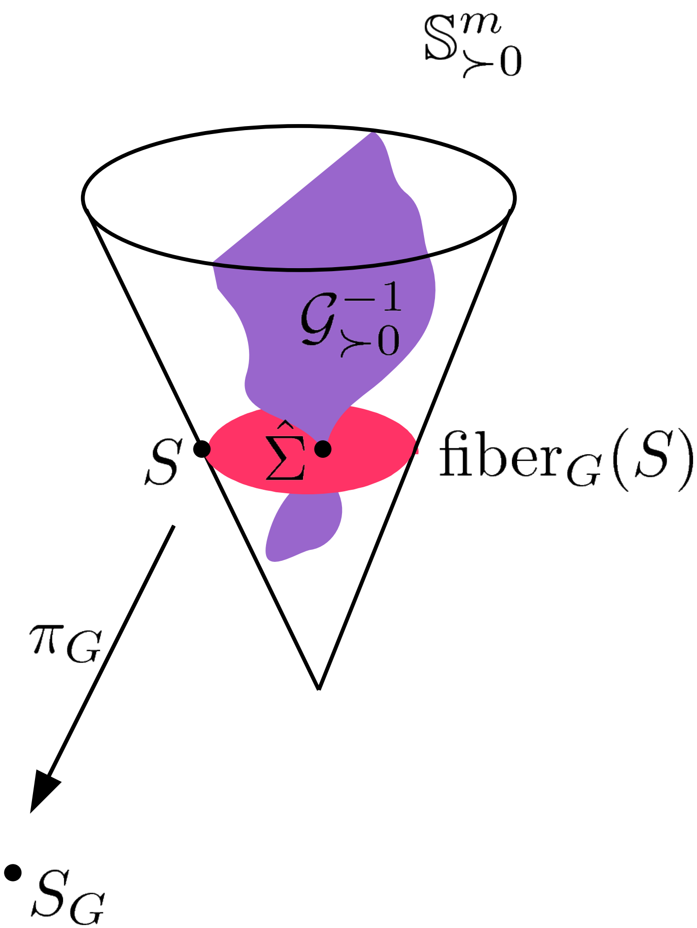
$$\hat{\Sigma}_{\alpha\beta} \;=\; S_{\alpha\beta}, \quad \forall(\alpha,\beta) \in E.$$

Geometry

Concentration matrices: K



Covariance matrices: Σ



Cones

- **Cone of concentration matrices:**

$$\mathcal{K}_G := \mathcal{G} \cap \mathbb{S}^m_{\succ 0}$$
$$= \left\{ (\lambda_{\alpha\beta})_{(\alpha,\beta)\in E} \in \mathbb{R}^E \ : \ \sum_{(\alpha,\beta)\in E} \lambda_{\alpha\beta} \mathbb{1}_{\alpha\beta} > 0 \right\}$$

- **Cone of sufficient statistics:**

$$\mathcal{C}_G := \pi_G(\mathbb{S}^m_{\succ 0})$$

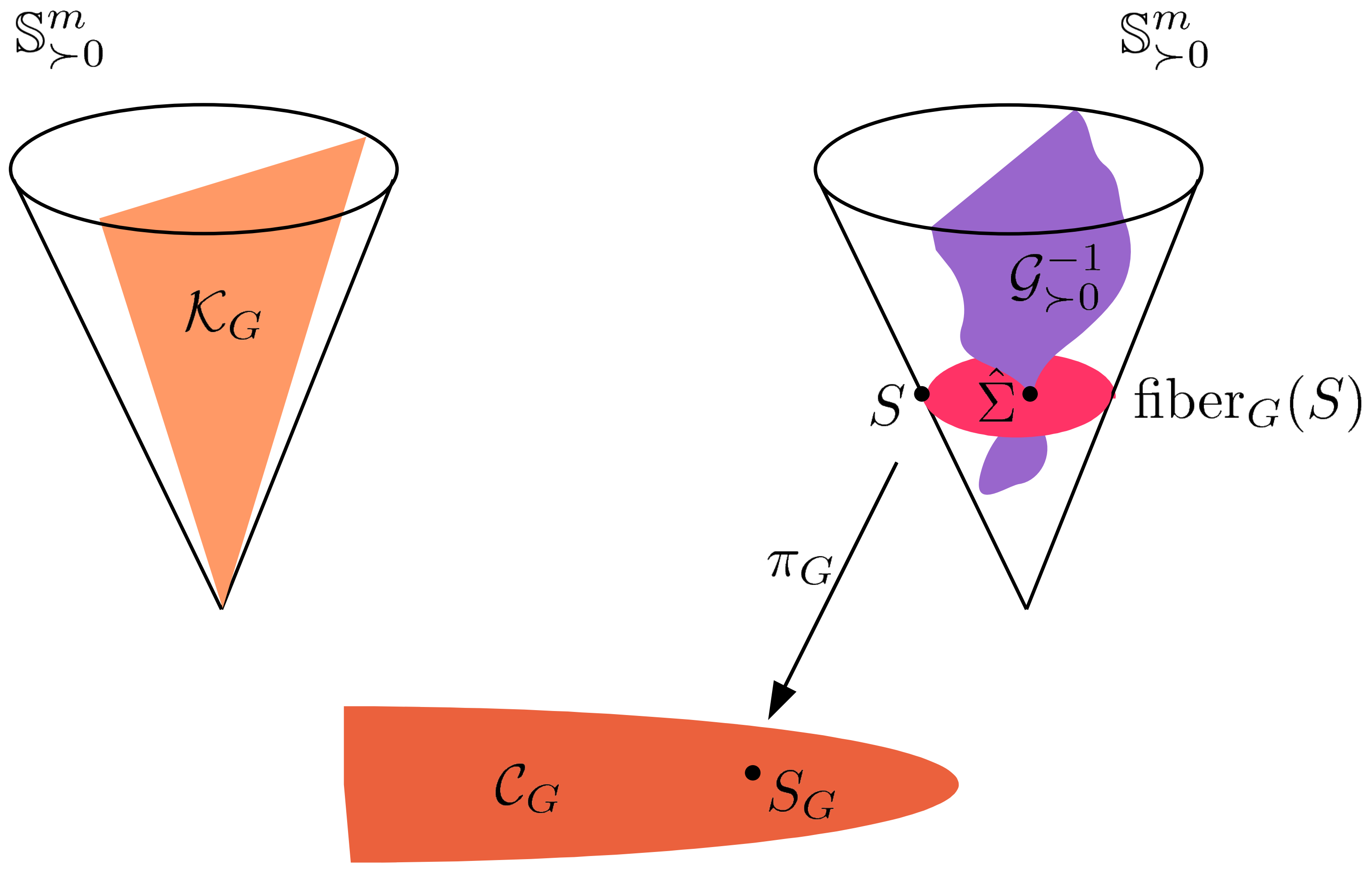
where $\pi_G : \mathbb{S}^m \to \mathbb{S}^m / \mathcal{G}^\perp$

respectively $\pi_G : \mathbb{S}^m \to \mathbb{R}^E, \quad S \mapsto S_G$

Geometry

Concentration matrices: K

Covariance matrices: Σ



Cones and maximum likelihood estimation

Theorem *(Sturmfels & U., 2010)*:

$\mathcal{C}_G$ is the convex dual to $\mathcal{K}_G$. Furthermore, $\overline{\mathcal{K}_G}$ and $\overline{\mathcal{C}_G}$ are closed convex cones which are dual to each other with

$$\overline{\mathcal{K}_G} = \mathcal{G} \cap \mathbb{S}^m_{\succeq 0} \qquad \text{and} \qquad \overline{\mathcal{C}_G} = \pi_G(\mathbb{S}^m_{\succeq 0}).$$

Theorem *(exponential families)*:

The map

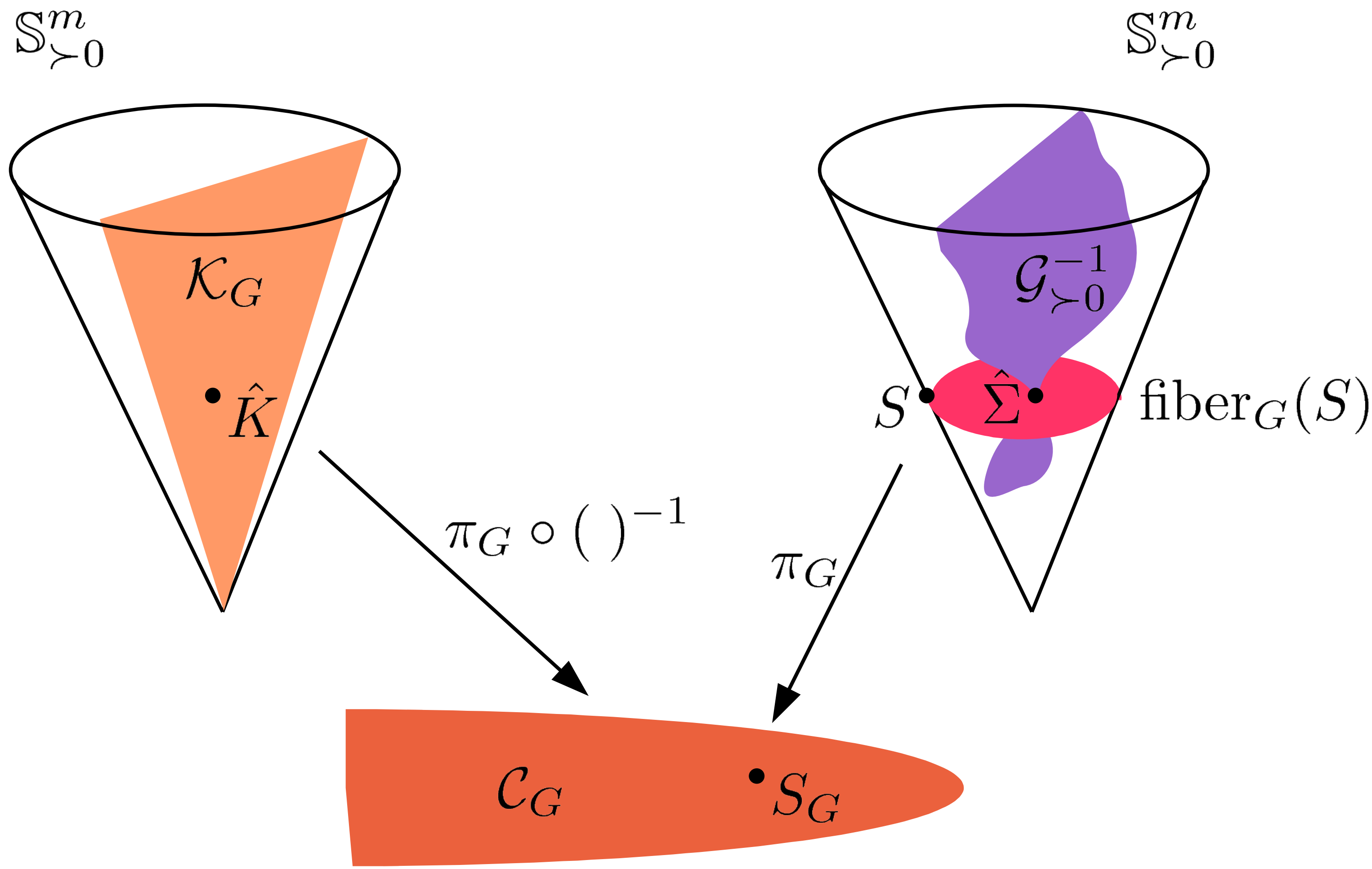
$$K \mapsto T = \pi_G(K^{-1})$$

is a homeomorphism between $\mathcal{K}_G$ and $\mathcal{C}_G$.

The inverse map $T \mapsto K$ takes the sufficient statistics to the MLE of the concentration matrix. Here, K^{-1} is the unique maximizer of the determinant over $\pi_G^{-1}(T) \cap \mathbb{S}^m_{\succ 0}$.

Geometry

Concentration matrices: K Covariance matrices: Σ



Existence of MLE: 2 Problems

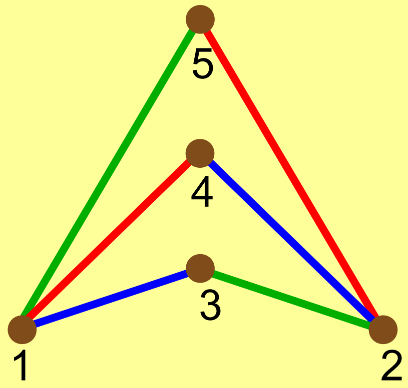
Given a graph G :

- Under what conditions on S_G does the MLE exist? (i.e. describe $\mathcal{C}_G$)
  - **"Cone" problem**

- Under what conditions on $(n, \ (X_1, \ldots, X_n))$ does the MLE exist?
  - **"Probability" problem**

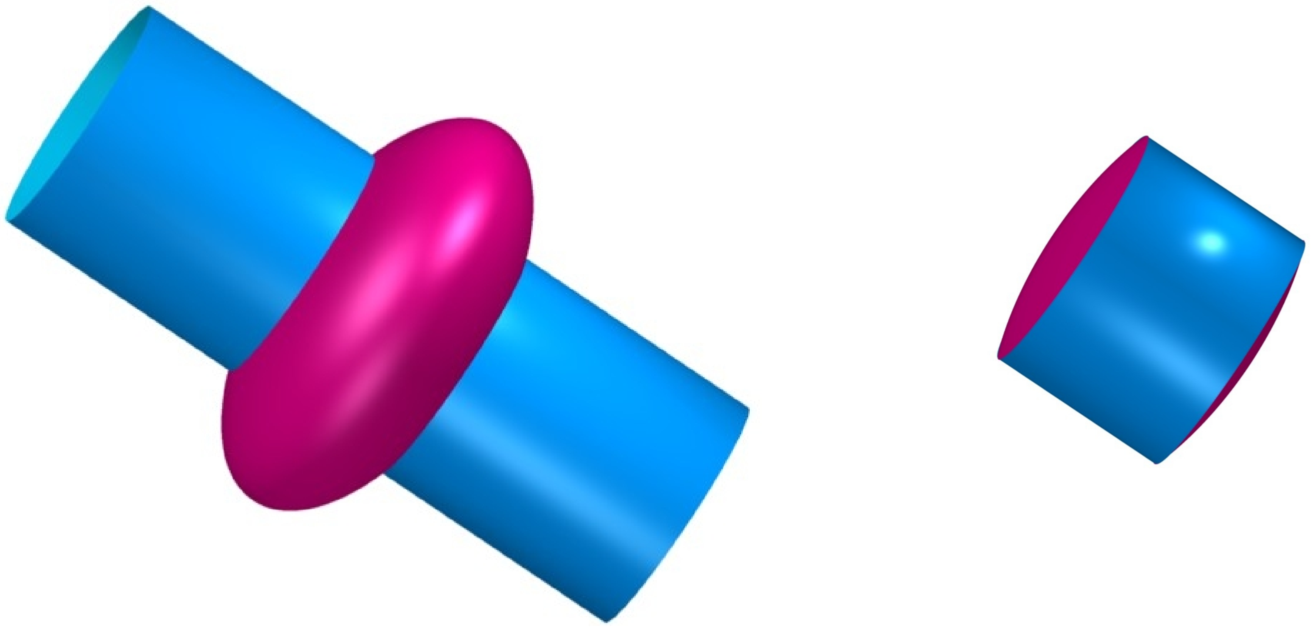
Problem 1: Example $K_{2,3}$

Example:



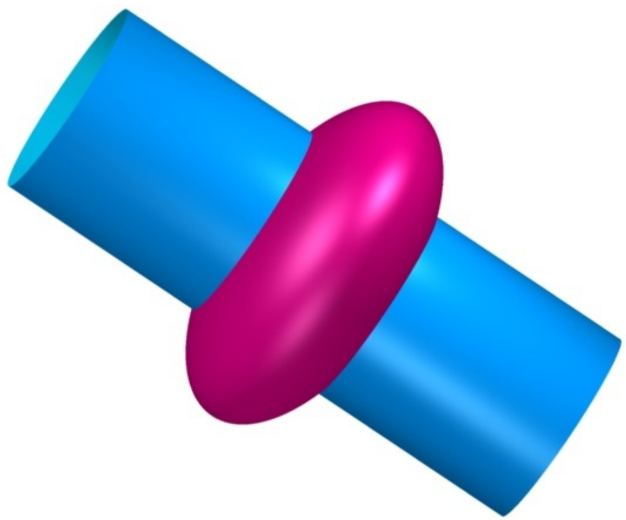
$$K = \begin{pmatrix} \lambda_1 & 0 & \lambda_2 & \lambda_3 & \lambda_4 \\ 0 & \lambda_1 & \lambda_4 & \lambda_2 & \lambda_3 \\ \lambda_2 & \lambda_4 & \lambda_1 & 0 & 0 \\ \lambda_3 & \lambda_2 & 0 & \lambda_1 & 0 \\ \lambda_4 & \lambda_3 & 0 & 0 & \lambda_1 \end{pmatrix}$$

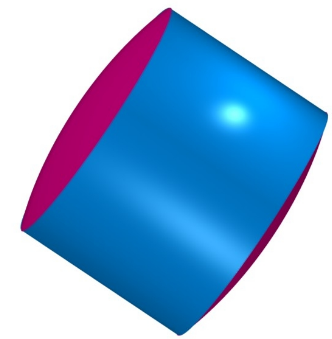
$\partial \mathcal{K}_G$:



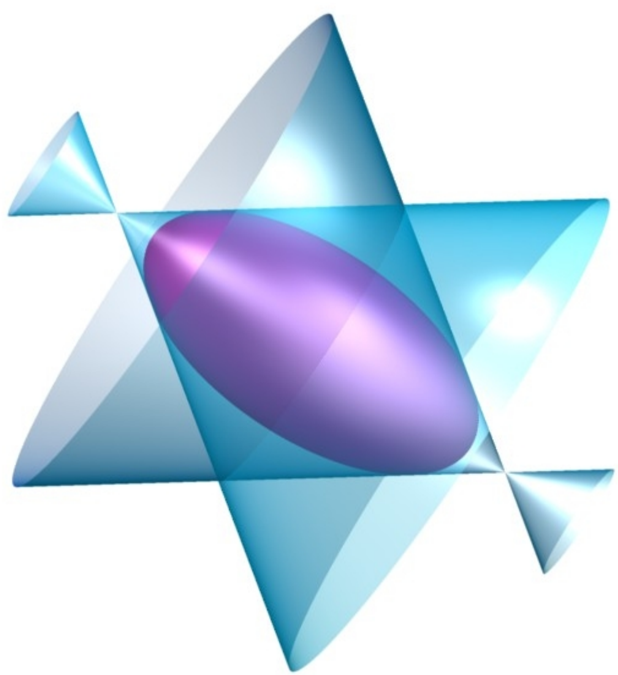
Problem 1: Example $K_{2,3}$

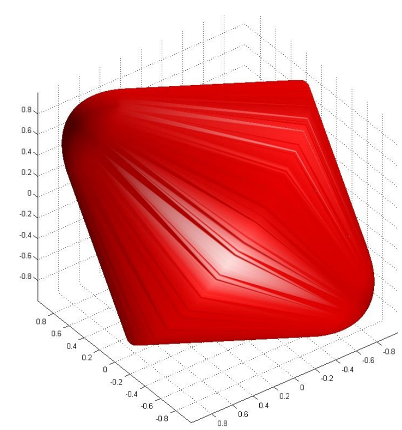
$\partial\mathcal{K}_G$:



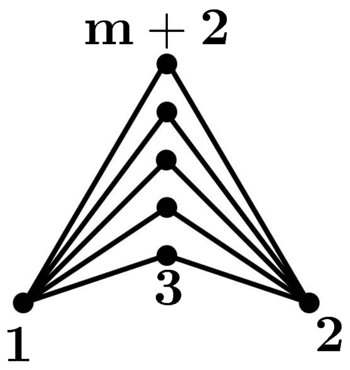


$\partial\mathcal{C}_G$:





Problem 1: $K_{2,m}$

$G =$ 

Corollary *(U.)*:

The MLE exists for S_G if and only if $\forall a \in \{1,2\},\ \ i,j \in \{3,\ldots,m\}$

$$2\arccos(S_{a,i}) \ < \ \sum_{b=1,2} \arccos(S_{b,i}) + \sum_{b=1,2} \arccos(S_{b,j})$$

$$< 2\pi + 2\arccos(S_{a,i})$$

Existence of MLE: 2 Problems

Given a graph G :

- Under what conditions on S_G does the MLE exist?

- Under what conditions on $(n, (X_1, \ldots, X_n))$ does the MLE exist?

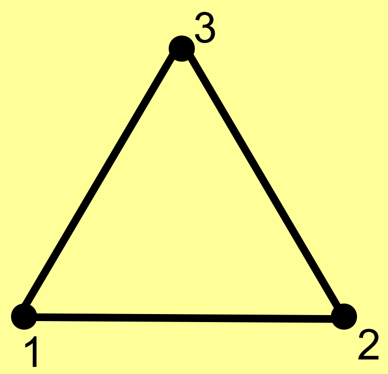
  And with what probability?

Probability of existence

Reminder: MLE exists $\Longleftrightarrow S_G \in \mathcal{C}_G$

$$\mathbb{P}_X(\text{MLE exists}) = \mathbb{P}_X(S_G \in \mathcal{C}_G)$$

Example:



$x_1, \ldots, x_n \sim \mathcal{N}_3(0, \Sigma)$

MLE exists $\Longleftrightarrow S \text{ PD} \Longleftrightarrow X \text{ full rank}$

$n < 3:$ MLE does not exist

$n \geq 3:$ MLE exists with probability 1

Probability of existence

Reminder: MLE exists $\Longleftrightarrow S_G \in \mathcal{C}_G$

$$\mathbb{P}_X(\text{MLE exists}) = \mathbb{P}_X(S_G \in \mathcal{C}_G)$$

Note: Existence of the MLE is **invariant** under

a) **Rescaling:**

$$X \to AX,\; S \to ASA, \quad \text{where } A \text{ is diagonal.}$$

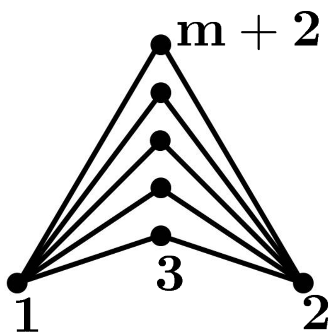
b) **Orthogonal transformation:**

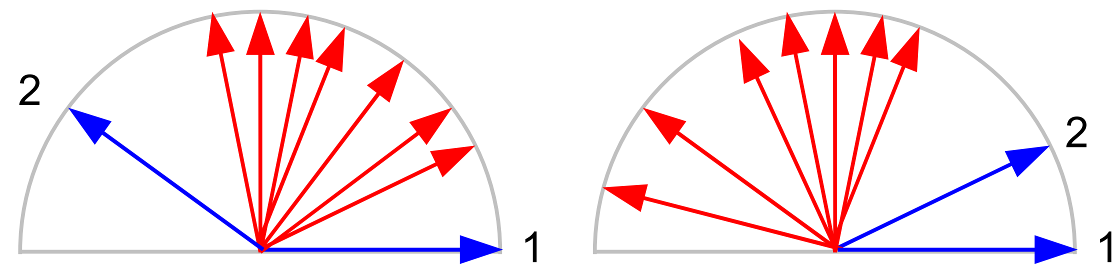
$$X \to XU,\; S \to XUU^T X^T = S, \quad \text{where } U \text{ is orthogonal.}$$

- Assume: $x_1, \ldots, x_m \in \mathbb{R}^n$ have length 1 and $x_1 = (1, 0, \ldots, 0)$.

Problem 2: $K_{2,m}$

$G =$





Theorem *(U.)*:

The MLE exists on $K_{2,m}$ with probability 1 for $n \geq 3$ and does not exist for $n = 1$.

For $n = 2$ let $x_1, \ldots, x_{m+2} \in \mathbb{R}^2$ ($x_1 = (1,0)$, $x_i = (\cos\omega_i, \sin\omega_i)$). The MLE exists if and only if $x_3, \ldots, x_{m+2}$ lie between x_1 and x_2 or $x_3, \ldots, x_{m+2}$ lie outside x_1 and x_2. This happens with prob. $\in (0, 1)$.

ML-degree of a graph

- The **maximum likelihood degree** of a statistical model is the number of complex solutions to the likelihood equations for generic data.

- **Generic data:** The number of solutions is a constant for all data, except possibly for a lower-dimensional subset of the data space.

ML-degree of a graph

Theorem *(Sturmfels & U., 2010)*:

G chordal if and only if

$$\text{ML-degree}(G) = 1.$$

Conjecture *(Drton, Sullivant & Sturmfels, 2009)*:

The ML-degree of an m-cycle C_m is given by

$$\text{ML-degree}(C_m) = (m-3)2^{m-2} + 1.$$

Theorem (*U.*):

The ML-degree of the bipartite graph $K_{2,m}$ is given by

$$\text{ML-degree}(K_{2,m}) = 2m + 1.$$

- Sturmfels & U.: Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry (AISM 62, 2010)
- U.: Maximum likelihood estimation in Gaussian graphical models (in progress)

Thank you!

- Malaspinas & U.: Detecting epistasis via Markov bases (arXiv:1006.4929)