Geometry of ML estimation in Gaussian graphical models

Caroline Uhler

Department of Statistics

UC Berkeley

ETH

December 17, 2010

Motivation

Current statistical applications:

- Number of variables >> Number of observations
- Example: Genetic networks

  Gene expression data of a few individuals to model interaction between large number of genes

➔ **Gaussian graphical models** widely used in this context

? Minimum number of observations for existence of MLE in Gaussian graphical model?

Gaussian graphical models

- $G = ([m], E)$ undirected graph with $(\alpha, \alpha) \in E \quad \forall \alpha \in [m]$.
- $X_1, \ldots, X_n \in \mathbb{R}^m$ i.i.d. sample from $\mathcal{N}_m(0, \Sigma)$
- $\Sigma \in \mathbb{S}^m_{\succ 0}$ covariance matrix on $[m]$ satisfying

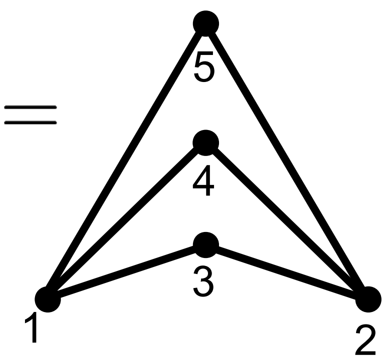
  $(\Sigma^{-1})_{\alpha,\beta} = 0, \quad \forall\, (\alpha, \beta) \notin E.$

- $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T \in \mathbb{S}^m_{\succeq 0}$ sample covariance matrix
- $S_G = (S_{\alpha,\beta} \mid (\alpha, \beta) \in E)$ sufficient statistics

- S_G **partial matrix** with entries at positions corresponding to edges E

Example $K_{2,3}$

- The Gaussian graphical model on $K_{2,3}$ consists of multivariate Gaussians with concentration matrix of the form

$G =$ 

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

- Given a sample covariance matrix S, the sufficient statistic is

$$S_G = \begin{pmatrix} s_{11} & ? & s_{13} & s_{14} & s_{15} \\ ? & s_{22} & s_{23} & s_{24} & s_{25} \\ s_{13} & s_{23} & s_{33} & ? & ? \\ s_{14} & s_{24} & ? & s_{44} & ? \\ s_{15} & s_{25} & ? & ? & s_{55} \end{pmatrix}$$

MLE, a special PD completion

Theorem *(Dempster, 1972)*:

In a Gaussian graphical model on G the MLE $\hat{\Sigma}$ exists, if and only if the partial sample covariance matrix S_G can be completed to a positive definite matrix.

Then the MLE $\hat{\Sigma}$ is the unique completion, whose inverse satisfies

$$(\hat{\Sigma}^{-1})_{\alpha,\beta} = 0, \;\; \forall \; (\alpha, \beta) \notin E.$$

- Existence of MLE in Gaussian graphical models is special **PD matrix completion problem** with rank constraint given by the number of observations.

PD matrix completion problems

Def: A graph G is **chordal** if every cycle of length ≥ 4 has a chord.

Key fact *(Grone et al. 1984)*

For a fixed graph G the following are equivalent:

i) Every partially pd matrix $M_G \in \mathbb{R}^E$ has a pd completion

ii) G chordal

q : maximal clique size of G

q^* : maximal clique size of a minimal chordal cover of G

Corollary:

If $n \geq q^*$ MLE exists with probability 1. If $n < q$ MLE does not exist.

? What happens in the gap $q \leq n < q^*$?

Outline

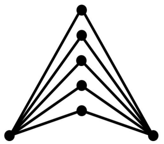
- Geometry of Dempster's theorem
  - Characterization of sufficient statistics for which MLE exists
  - Characterization of minimal number of observations needed

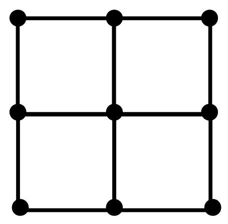
- Examples:

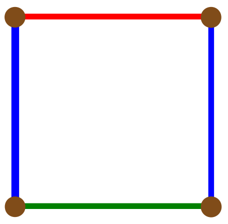
  - Bipartite graphs $K_{2,m}$

  - 3-by-3 grid

  - Colored 4-cycle

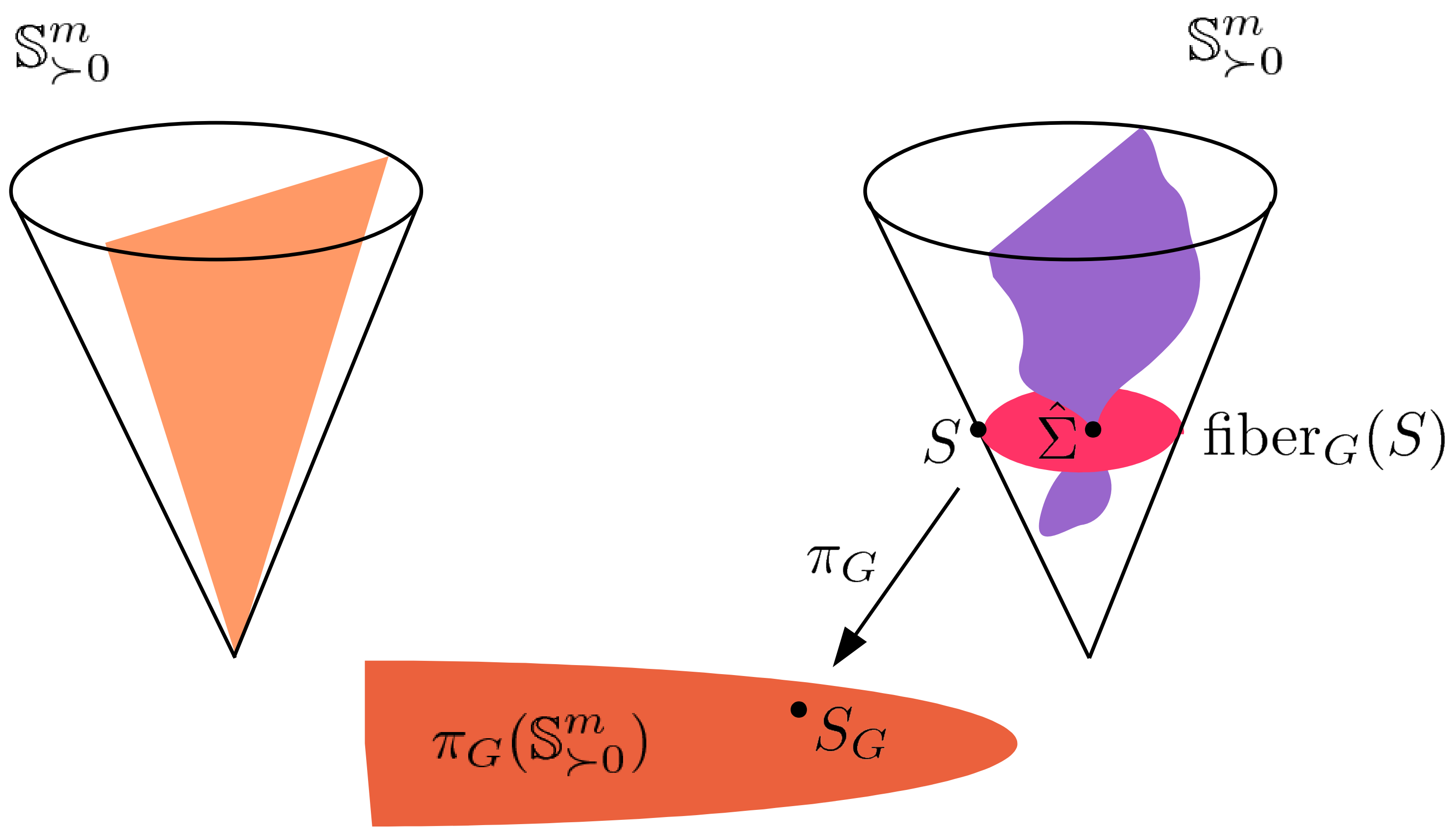






Geometry of Dempster's theorem

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



$$\mathrm{fiber}_G(S) := \{\Sigma \in \mathbb{S}^m_{\succ 0} \mid \Sigma_G = S_G\}$$

Cones

Def: $\mathcal{C} \subset \mathbb{R}^k$ is a **convex cone** iff

$$ax + by \in \mathcal{C} \quad \text{for all} \quad a, b \geq 0, \quad x, y \in \mathcal{C}.$$

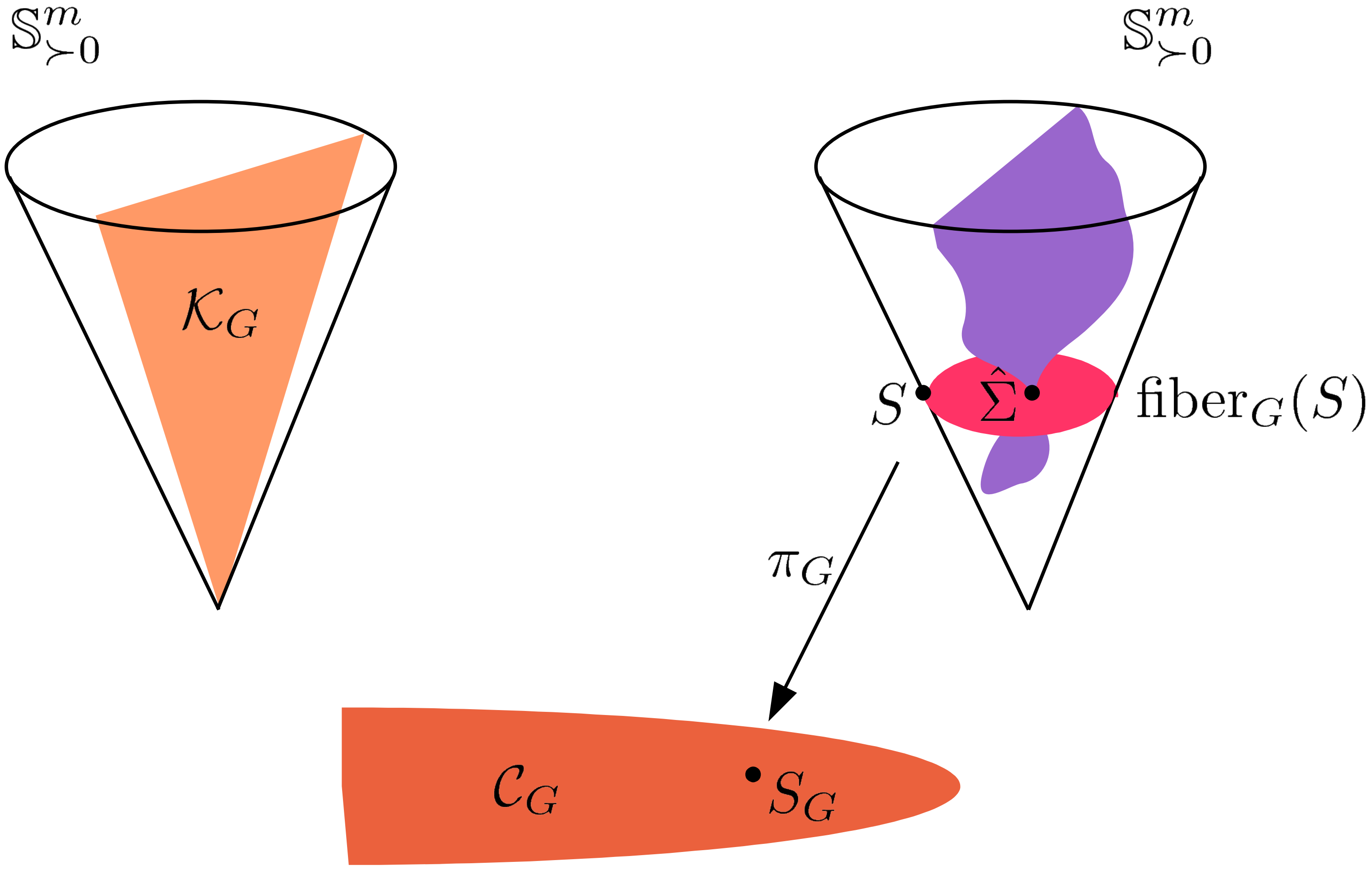
Ex: $\mathbb{R}^m_+, \quad \mathbb{S}^m_{\succeq 0}$

- **Cone of concentration matrices:** $\mathcal{K}_G \ := \ \mathcal{G} \cap \mathbb{S}^m_{\succ 0}$
- **Cone of sufficient statistics:** $\mathcal{C}_G \ := \ \pi_G(\mathbb{S}^m_{\succ 0})$

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

Geometry of ML estimation

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



Cones and maximum likelihood estimation

Def: Let $\mathcal{C}$ be a convex cone. The **dual cone** is

$$C^* = \{w \mid \langle v, w\rangle \geq 0 \quad \text{for all } v \in \mathcal{C}\}.$$

Ex: $\mathbb{R}^m_+$, $\mathbb{S}^m_{\succeq 0}$ are self-dual

Theorem *(Sturmfels & U., 1972)*:

$\mathcal{C}_G$ 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

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

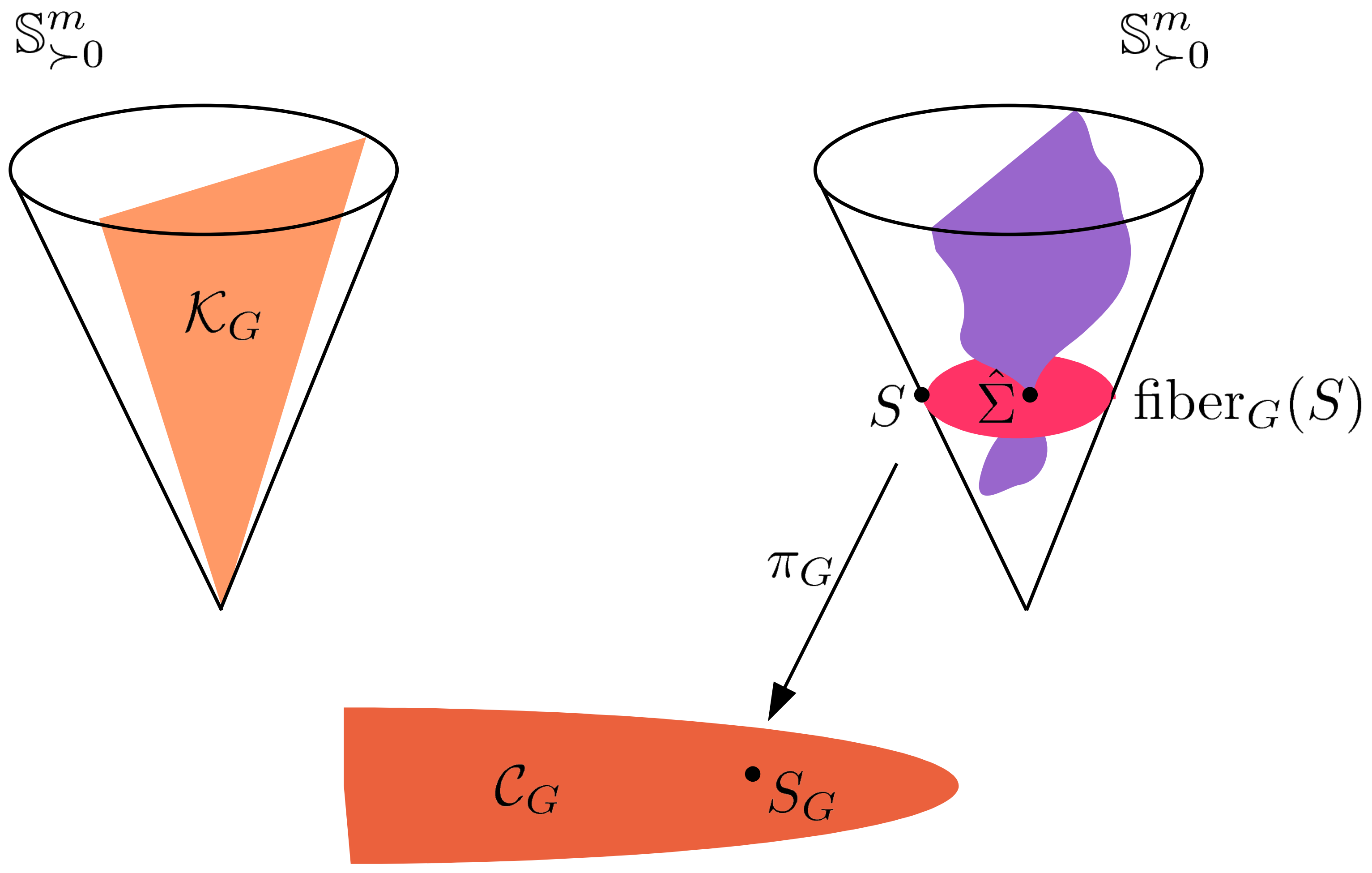
Theorem:

If the MLE exists, it is defined as the unique solution to:

maximize	$\log\det K - \mathrm{tr}(KS)$	minimize	$-\log\det\Sigma$
subject to	$K \in \mathcal{K}_G$	subject to	$\Sigma \in \mathrm{fiber}_G(S)$

Geometry of ML estimation

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

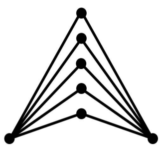


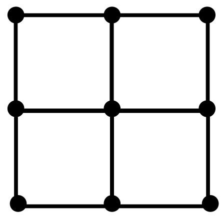
Outline

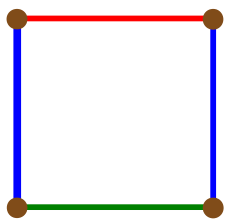
- Geometry of Dempster's theorem
  - Characterization of sufficient statistics for which MLE exists
  - Characterization of minimal number of observations needed

- Examples:

  - Bipartite graphs $K_{2,m}$
  - 3-by-3 grid
  - Colored 4-cycle

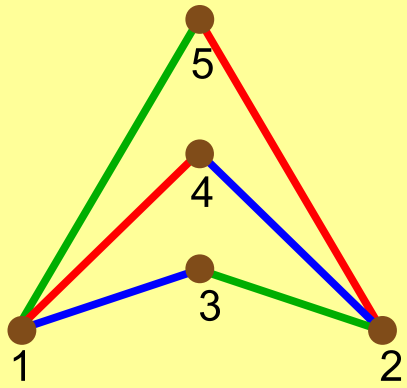






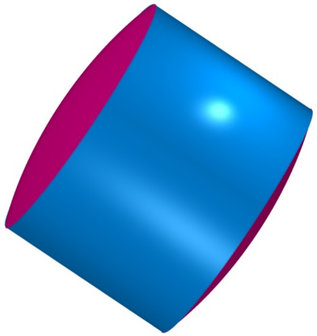
Example $K_{2,3}$

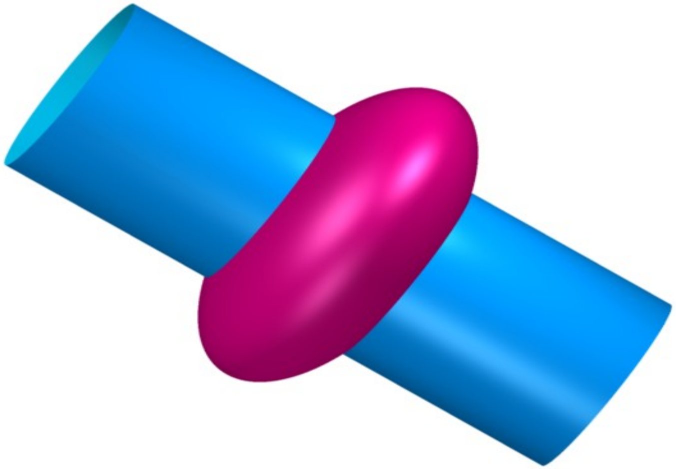
Example:



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

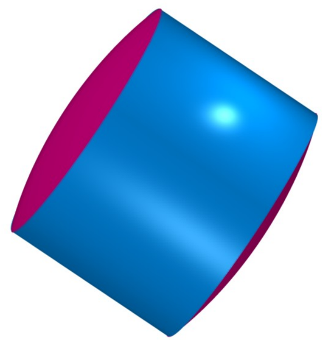
$\mathcal{K}_G$:

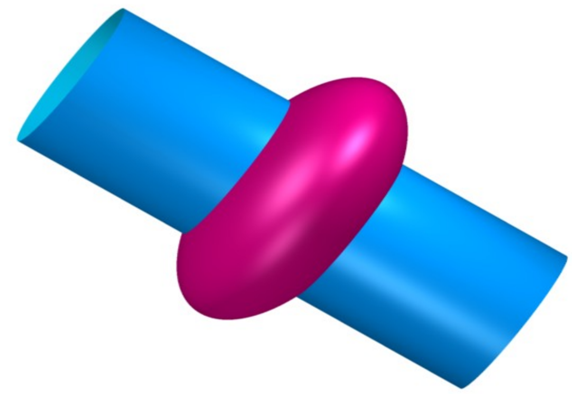




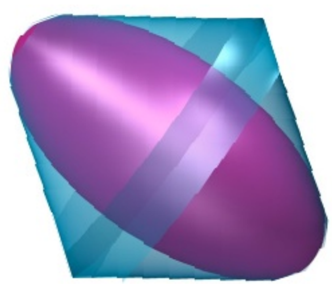
Example $K_{2,3}$

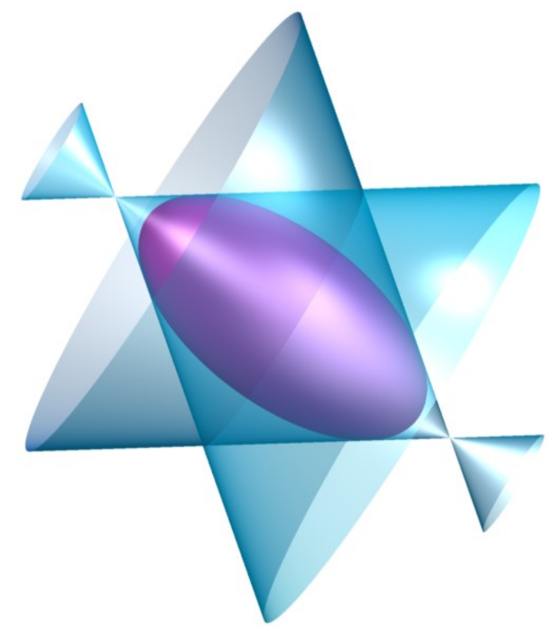
$\mathcal{K}_G$:



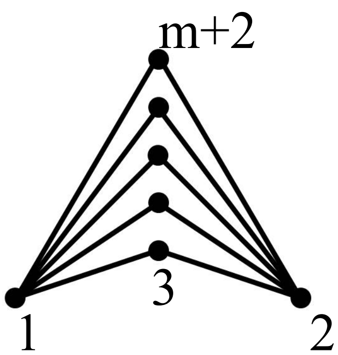


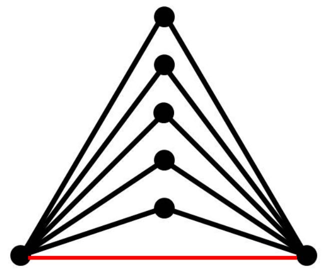
$\mathcal{C}_G$:

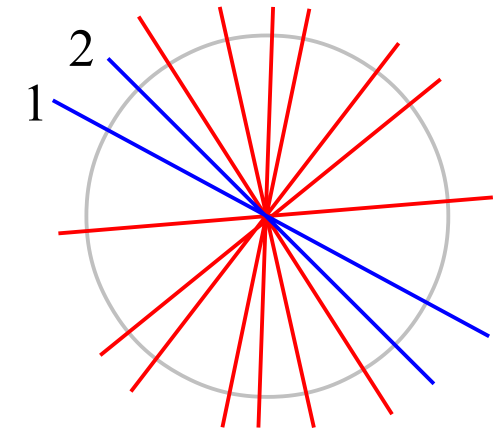




$K_{2,m}$

$G =$ 





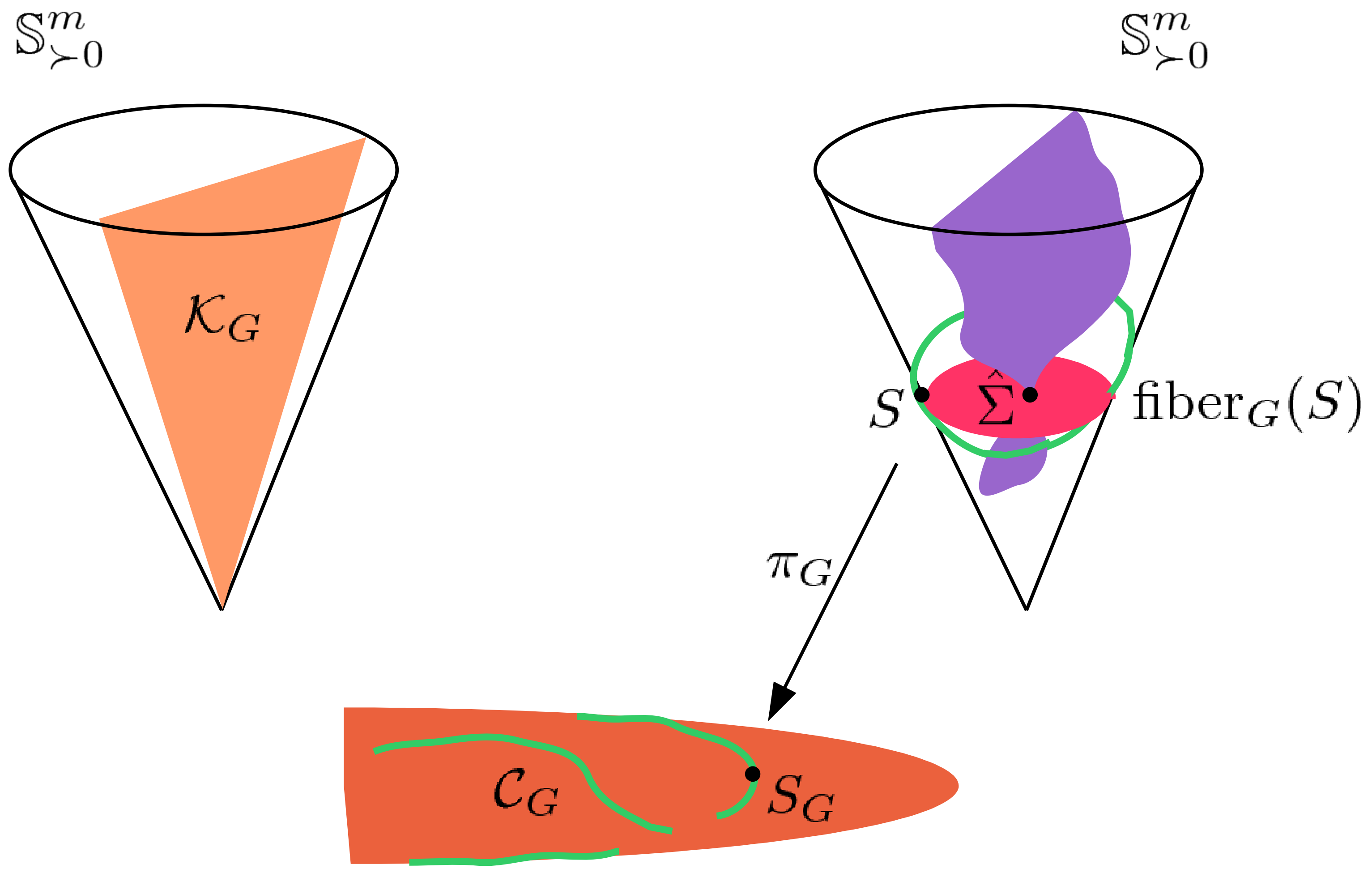
Theorem *(U. 2010)*:

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 $x_1, \ldots, x_{m+2} \in \mathbb{R}^2$. The MLE exists if and only if the lines corresponding to x_1 and x_2 are neighbors. This happens with probability $\in (0, 1)$.

Geometry of ML estimation

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



Sufficient condition for existence of MLE

Elimination Criterion *(U. 2010)*:

Let I_n be the ideal of $(n+1) \times (n+1)$ minors of a symmetric $m \times m$ matrix of unknowns S. Let $I_{G,n}$ be the elimination ideal obtained from I_n by eliminating all unknowns corresponding to non-edges in the graph. If

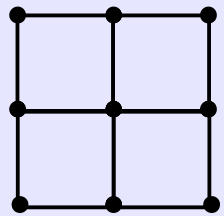
$$I_{G,n} = 0,$$

then the MLE exists with probability 1 for n observations.

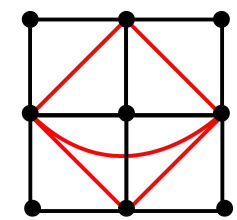
- I_n corresponds to all symmetric matrices of rank $\leq n$
- Elimination corresponds to projection onto $\mathcal{C}_G$
- $I_{G,n} = 0$ means that the projection is full-dimensional

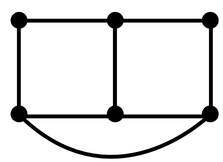
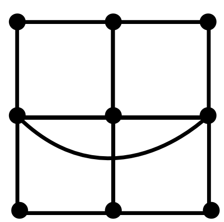
3-by-3 grid

Theorem (U. 2010):

The MLE on $G =$  exists with probability 1 for $n \geq 3$.

Solves Steffen Lauritzen's question



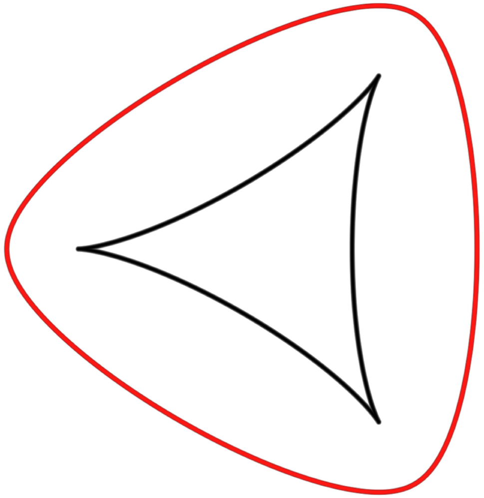
- $q = 2$ and $q^* = 4$
- For $\mathcal{G} =$  $I_{\mathcal{G},3} = 0$
- For $\mathcal{G} =$  MLE exists with probability 1 for $n \geq 3$
- For $n = 2$ sufficient conditions for existence of MLE still open

What if $I_{G,n}$ is not the zero ideal?

If $I_{G,n} \neq 0$, careful analysis of corresponding components is needed.

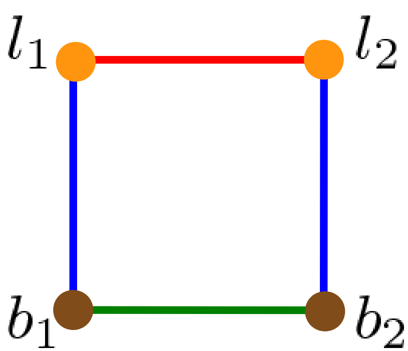
! Even if $I_{G,n}$ corresponds to a component of the algebraic boundary of $\mathcal{C}_G$, the MLE might still exist.

! The algebraic boundary of $\mathcal{C}_G$ might intersect the interior of $\mathcal{C}_G$.



Frets' heads *(Mardia, Kent, and Bibby, 1979)*

- Heredity study of head dimensions
- Length and breadth of heads of 25 pairs of first and second sons
- Data supports model, where joint distribution is unaltered if two sons are interchanged



$$K = \begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$$

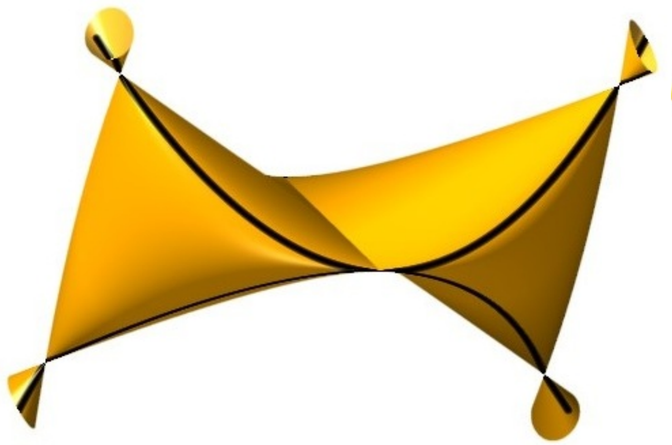
- Algebraic boundary of $\mathcal{C}_G$ defined by

$$(t_1 - t_3)(t_1 + t_3)(t_2 - t_5)(t_2 + t_5)(4t_2^2t_3^2 - 4t_1t_2t_4^2 + t_4^4 + 8t_1t_2t_3t_5 - 4t_3t_4^2t_5 - 4t_3t_4^2t_5 + 4t_1^2t_5^2)$$

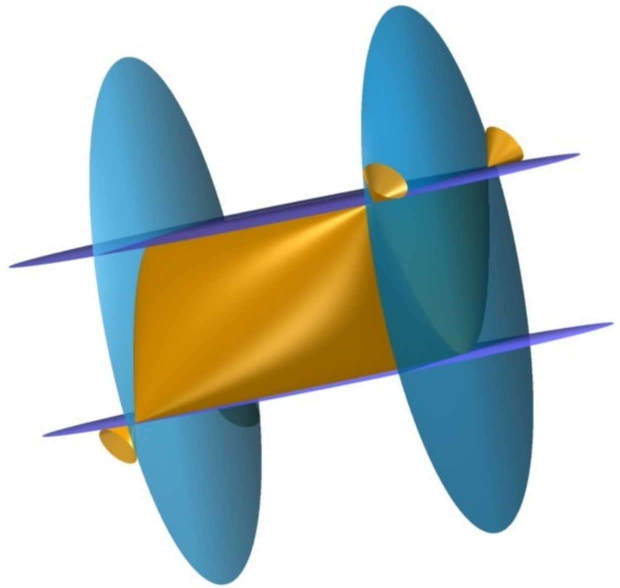
- $I_{G,1} = \langle 4t_2^2t_3^2 - 4t_1t_2t_4^2 + t_4^4 + 8t_1t_2t_3t_5 - 4t_3t_4^2t_5 - 4t_3t_4^2t_5 + 4t_1^2t_5^2 \rangle$

Frets' heads *(Mardia, Kent, and Bibby, 1979)*

$n = 1:$



$\mathcal{C}_G:$



- MLE exists for one observation if and only if sufficient statistics lie on "triangles" of bow tie

- MLE exists if and only if $\quad l_1 > l_2$ and $b_1 < b_2$

  or $\quad l_2 > l_1$ and $b_2 < b_1$

Conclusions and future directions

- Geometry of ML estimation in Gaussian graphical models is RICH
- In principle we can answer question about minimal number of observations for existence of MLE in any graph
- In practice only for small graphs

Ultimate goal: Application to huge gene association networks

- Use small graphs as bricks to build larger graphs
- Study asymptotics of ML estimation in Gaussian graphical models

Ex: $\mathbb{P}\left(\text{MLE exists},\ n=2,\ \Sigma_m^{\text{true}},\ G_m^{\text{assumed}} = \text{cycle}\right) \xrightarrow{m\to\infty} 1$

if condition number of Σ_m^{true} is $< \dfrac{m^{1-\frac{1}{2m}}}{e}$ for all m

- Sturmfels & U.: Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry (AISM 62, 2010)
- U.: Geometry of maximum likelihood estimation in Gaussian graphical models (on the arXiv since Tuesday)
- Chandrasekaran, Shah, U. Asymptotics of maximum likelihood estimation in Gaussian cycles (in progress)

Thank you!