Geometry of ML estimation in Gaussian graphical models

Caroline Uhler

Institute for Mathematics and its Applications

University of Minnesota

SIAM Conference on Applied Algebraic Geometry

October 8, 2011

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 (maximum likelihood estimator) in Gaussian graphical model?

Gaussian graphical models

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

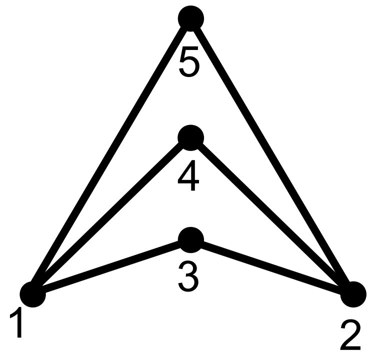
  $(\Sigma^{-1})_{\alpha,\beta} = 0, \ \ \forall\, (\alpha, \beta) \notin E.$

- $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T \in \mathbb{S}^m_{\succeq 0}$ 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 satisfying

$G =$ 

$$\Sigma^{-1} = \begin{pmatrix} \lambda_{11} & 0 & \lambda_{13} & \lambda_{14} & \lambda_{15} \\ 0 & \lambda_{22} & \lambda_{23} & \lambda_{24} & \lambda_{25} \\ \lambda_{13} & \lambda_{23} & \lambda_{33} & 0 & 0 \\ \lambda_{14} & \lambda_{24} & 0 & \lambda_{44} & 0 \\ \lambda_{15} & \lambda_{25} & 0 & 0 & \lambda_{55} \end{pmatrix}$$

- 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)*:

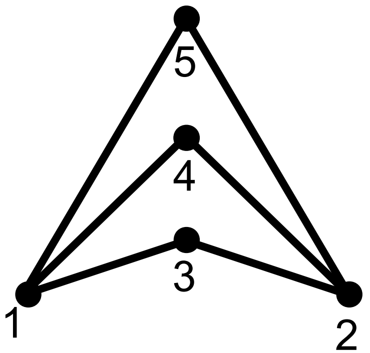
In a Gaussian graphical model on G the MLE $\hat{\Sigma}$ exists, if and only if the partial sample covariance matrix S_G can be completed to a positive definite matrix.

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

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

Example $K_{2,3}$

- The Gaussian graphical model on $K_{2,3}$ consists of multivariate Gaussians satisfying

$G =$ 

$$\Sigma^{-1} = \begin{pmatrix} \lambda_{11} & 0 & \lambda_{13} & \lambda_{14} & \lambda_{15} \\ 0 & \lambda_{22} & \lambda_{23} & \lambda_{24} & \lambda_{25} \\ \lambda_{13} & \lambda_{23} & \lambda_{33} & 0 & 0 \\ \lambda_{14} & \lambda_{24} & 0 & \lambda_{44} & 0 \\ \lambda_{15} & \lambda_{25} & 0 & 0 & \lambda_{55} \end{pmatrix}$$

- 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

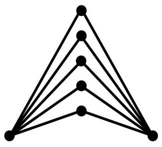
$$(\hat{\Sigma}^{-1})_{\alpha,\beta} = 0, \;\; \forall \; (\alpha, \beta) \notin E.$$

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

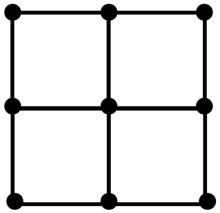
Outline

- Geometry of Dempster's theorem
  - Characterization of minimal number of observations needed

- Examples
  - Bipartite graphs $K_{2,m}$

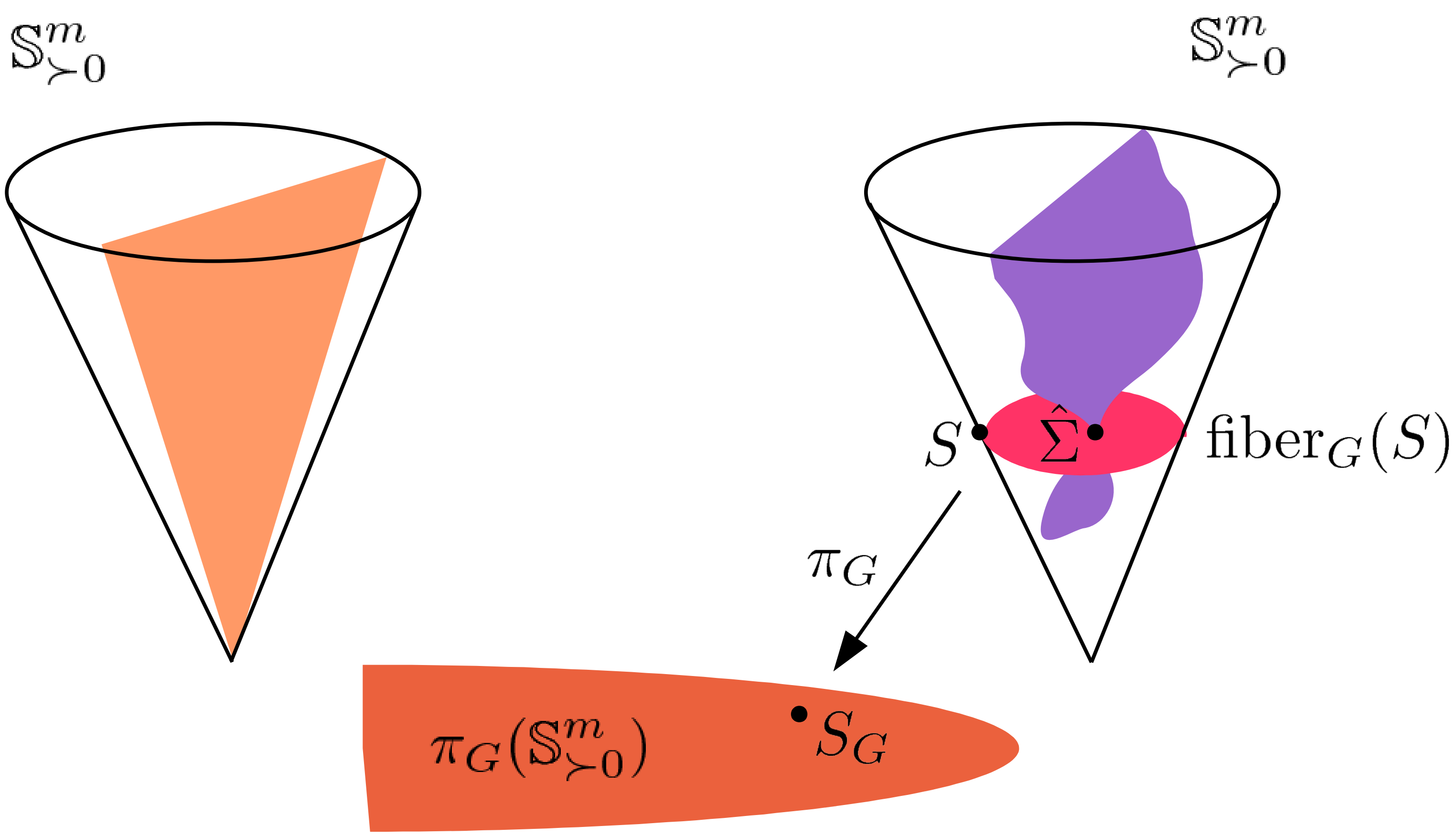


  - 3-by-3 grid



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** if

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

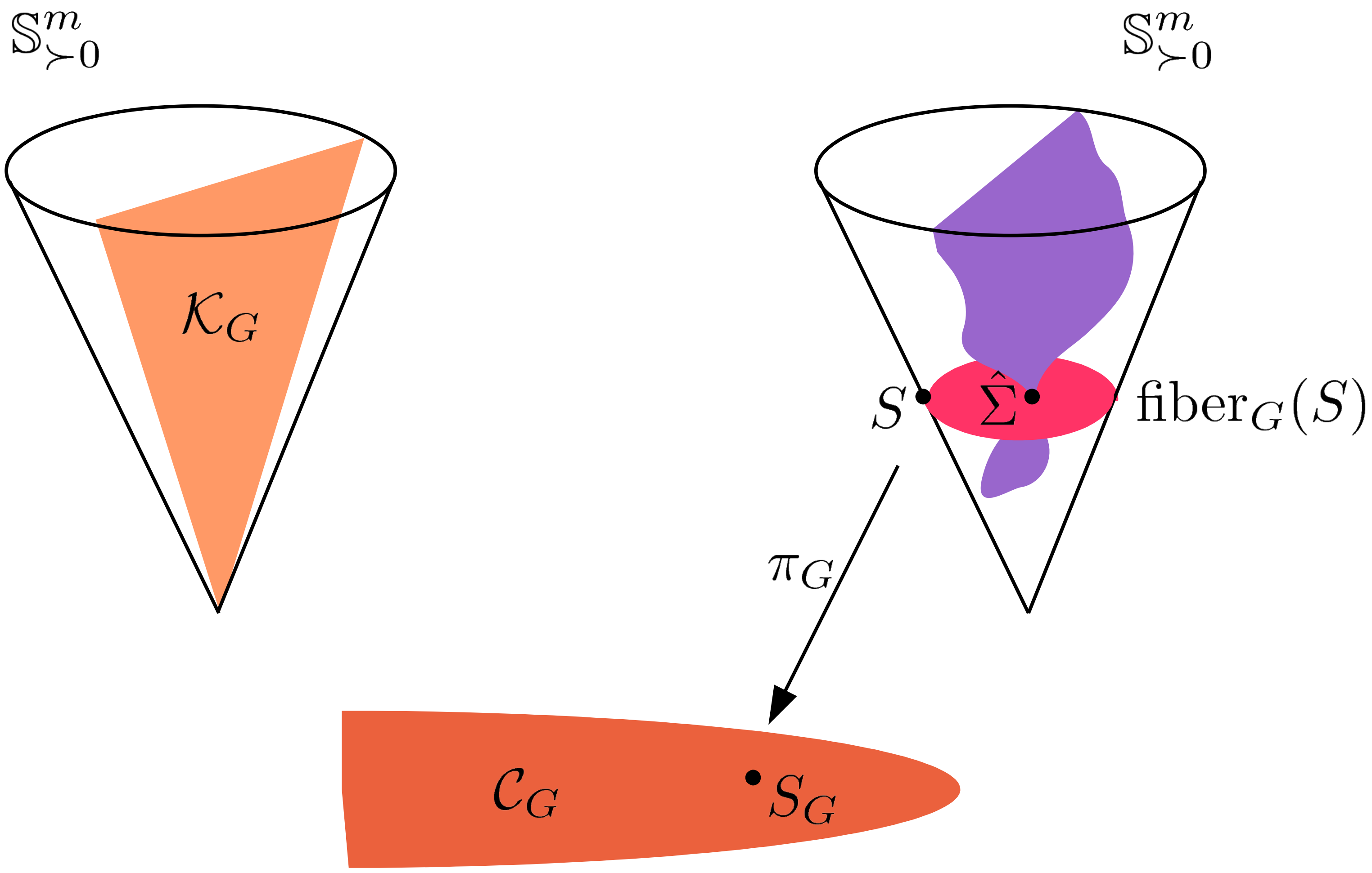
Ex: $\mathbb{R}^m_+, \ \ \mathbb{S}^m_{\succeq 0}$

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

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

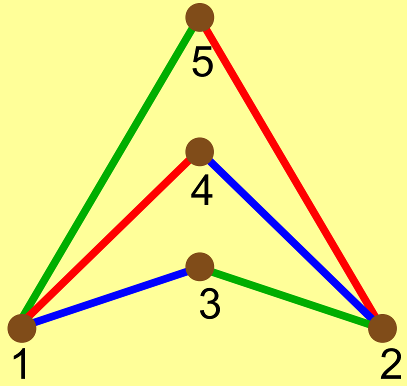
Theorem *(Sturmfels & U., 2010)*:
$\mathcal{C}_G$ is the convex dual to $\mathcal{K}_G$. Furthermore, $\overline{\mathcal{K}_G}$ and $\overline{\mathcal{C}_G}$ are closed convex cones which are dual to each other with

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

- Used this duality to find an algebraic characterization of the cone of sufficient statistics

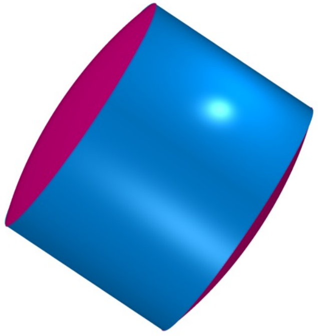
Example $K_{2,3}$

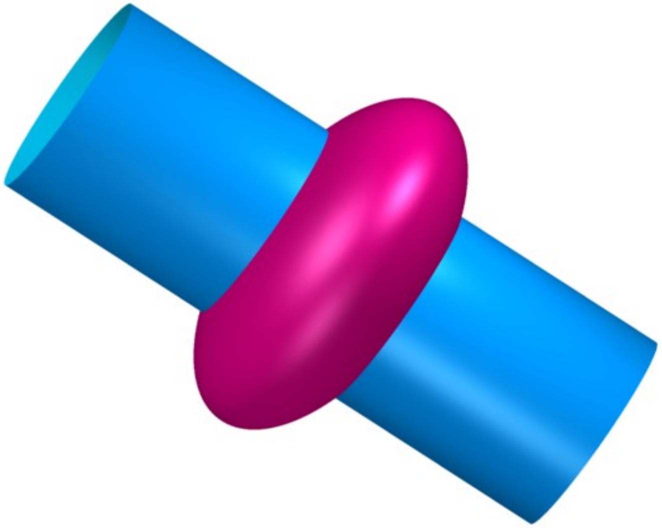
Example:



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

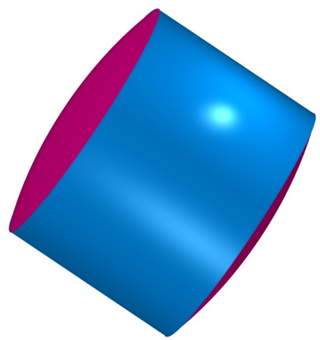
$\mathcal{K}_G$:

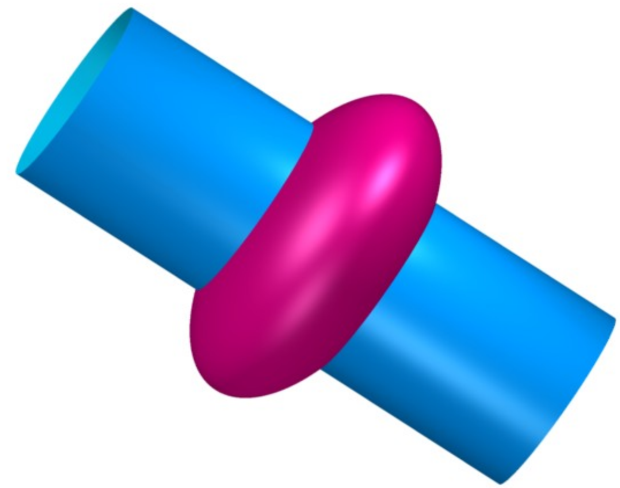




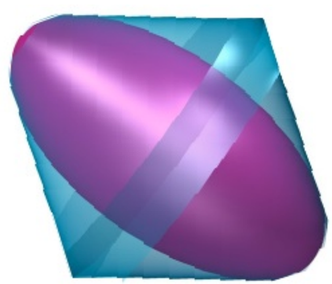
Example $K_{2,3}$

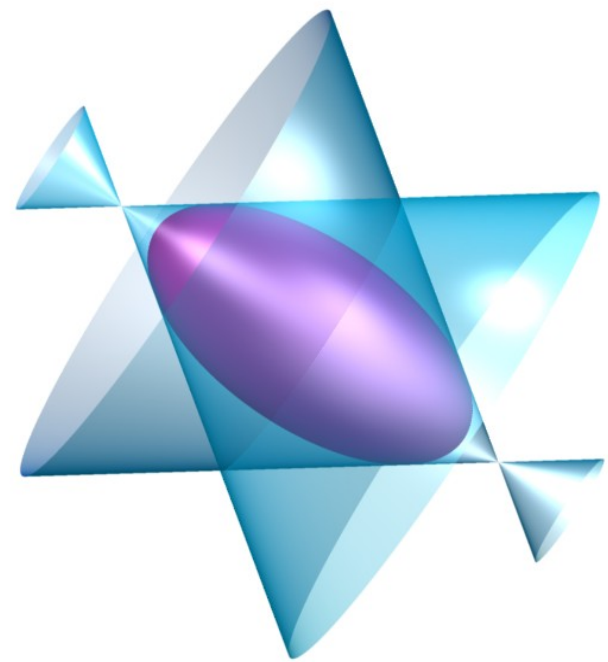
$\mathcal{K}_G$:





$\mathcal{C}_G$:

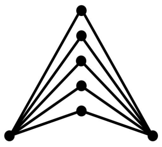




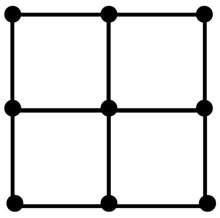
Outline

- Geometry of Dempster's theorem
  - Characterization of minimal number of observations needed

- Examples
  - Bipartite graphs $K_{2,m}$

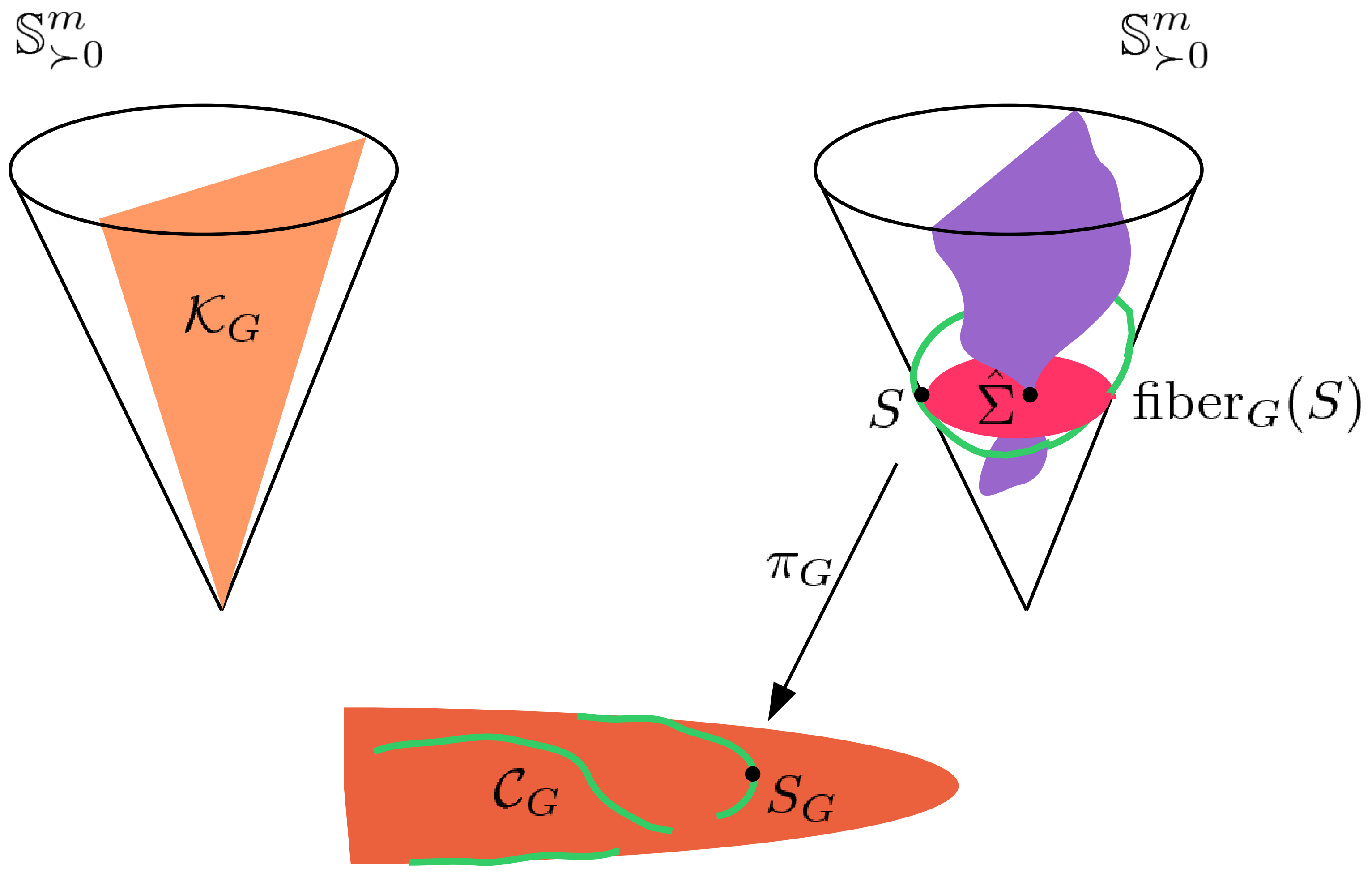


  - 3-by-3 grid



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

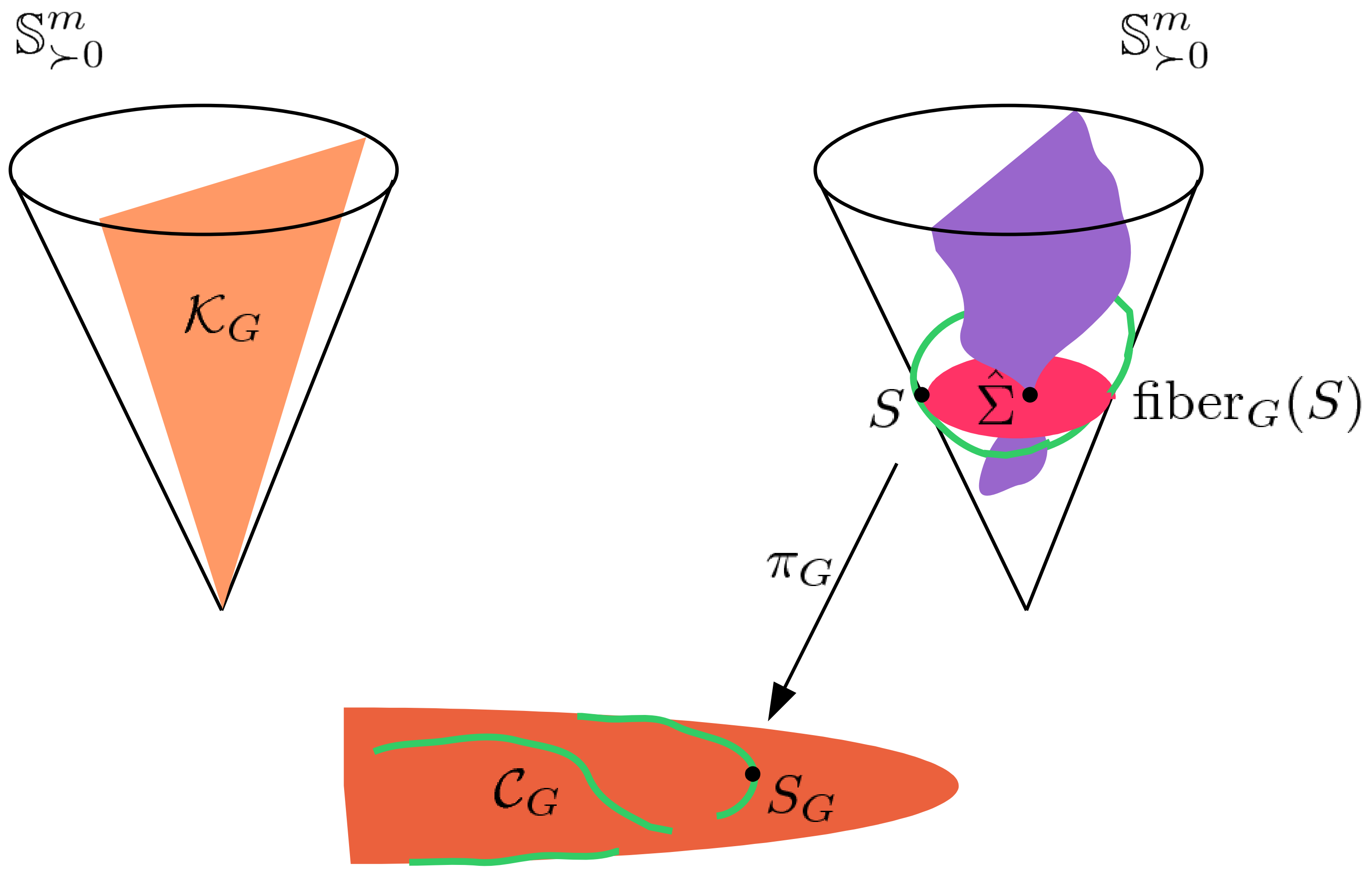
$$I_{G,n} = 0,$$

then the MLE exists with probability 1 for n observations.

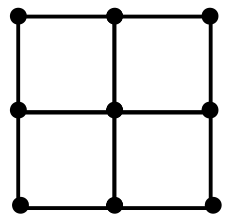
- I_n corresponds to all symmetric matrices of rank $\leq n$
- Elimination corresponds to projection onto $\mathcal{C}_G$
- $I_{G,n} = 0$ means that the projection is full-dimensional

Geometry of ML estimation

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



3-by-3 grid

$G =$ 

Theorem (U. 2010):

The MLE on the 3-by-3 grid exists with probability 1 if $n \geq 3$ and does not exist if $n = 1$. For $n = 2$ the MLE exists with probability $\in (0, 1)$.

Solves Steffen Lauritzen's open problem

- Sturmfels & U.: Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry (AISM 62, 2010)
- U.: Geometry of maximum likelihood estimation in Gaussian graphical models (submitted, arXiv:1012.2643v1)

Thank you!

- I am looking for a postdoc starting in September 2012 at the Institute of Science and Technology Austria.