Multivariate Gaussians, & Convex Algebraic Geometry

Caroline Uhler

Department of Statistics

UC Berkeley

(joint work with Bernd Sturmfels)

Second de Brùn Workshop

July 1, 2009

Outline

- Background and setup:

  - Multivariate Gaussians and linear concentration models

  - Maximum likelihood estimation

  - Convex algebraic geometry

- Guiding question

- 2 classes of linear concentration models

  - Gaussian graphical models

  - Colored Gaussian graphical models

Linear concentration models

- $\mathcal{N}_{\mathbf{m}}(\mathbf{0}, \mathbf{\Sigma})$:
  - $\Sigma \in \mathbb{S}^m_{\succ 0}$ covariance matrix
  - $\mathcal{L} := \langle K_1, \ldots, K_d \rangle$ linear subspace of $\mathbb{S}^m$
  - $K := \Sigma^{-1} \in \mathbb{S}^m_{\succ 0}$ concentration matrix with $K \in \mathcal{L}$

- **Linear concentration model:**

$$\mathcal{L}^{-1}_{\succ 0} := \left\{ \Sigma \in \mathbb{S}^m_{\succ 0} \ : \ \Sigma^{-1} \in \mathcal{L} \right\}$$

- **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
  - $\langle S, K_j \rangle,\ j = 1, \ldots d$ sufficient statistics

Example

$$K_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad K_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad K_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

- The linear concentration model consists of all multivariate Gaussians whose concentration matrix is of the form

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

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

$$t_1 = \mathrm{trace}(S), \quad t_2 = 2s_{12}, \quad t_3 = 2s_{13}, \quad t_4 = 2s_{23}.$$

Maximum likelihood estimation

- Log-likelihood function:

$$\log\det(K) - \langle S, K\rangle \;=\; \log\det\left(\sum_{j=1}^{d}\lambda_j K_j\right) - \sum_{j=1}^{d}\lambda_j\langle S, K_j\rangle$$

Theorem *(exponential families)*:

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

$$\mathrm{fiber}_{\mathcal{L}}(S) := \{\Sigma \in \mathbb{S}^m_{\succ 0} \,:\, \langle \Sigma, K_j\rangle = \langle S, K_j\rangle,\; j = 1,\dots,d\} \neq \emptyset.$$

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

$$\langle \hat{\Sigma}, K_j\rangle = \langle S, K_j\rangle \qquad \text{for} \quad j = 1,\dots,d.$$

Main problems

- Under what conditions on $(\mathcal{L},\ S)$ does the MLE exist?
- Under what conditions on $(\mathcal{L},\ n,\ (X_1, \ldots, X_n))$ does the MLE exist?
- In this talk we'll study the first problem.

Cones

- **Cone of concentration matrices:**

$$\begin{aligned} \mathcal{K}_{\mathcal{L}} \; &:= \; \mathcal{L} \cap \mathbb{S}^m_{\succ 0} \\ &= \Big\{ (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d \; : \; \textstyle\sum_{i=1}^d \lambda_i K_i \succ 0 \Big\} \end{aligned}$$

- **Cone of sufficient statistics:**

$$\mathcal{C}_{\mathcal{L}} \; := \; \pi_{\mathcal{L}}(\mathbb{S}^m_{\succ 0})$$

where $\pi_{\mathcal{L}} : \mathbb{S}^m \to \mathbb{S}^m / \mathcal{L}^{\perp}$

respectively $\pi_{\mathcal{L}} : \mathbb{S}^m \to \mathbb{R}^d, \quad S \mapsto (\langle S, K_1 \rangle, \ldots, \langle S, K_d \rangle)$

Cones and statistical theory

Theorem:

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

$$\overline{\mathcal{K}_{\mathcal{L}}} = \mathcal{L} \cap \mathbb{S}^m_{\succeq 0} \qquad \text{and} \qquad \overline{\mathcal{C}_{\mathcal{L}}} = \pi_{\mathcal{L}}(\mathbb{S}^m_{\succeq 0}).$$

Theorem *(exponential families)***:**

The map

$$K \mapsto T = \pi_{\mathcal{L}}(K^{-1})$$

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

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 $\pi_{\mathcal{L}}^{-1}(T) \cap \mathbb{S}^m_{\succ 0}$.

Guiding question

❓ Under what conditions on $(\mathcal{L},\ S)$ does the MLE exist?

➡ Pass to $\mathbb{C}$ to make problem **easier**.

Guiding question:

Determine hypersurface $\partial \mathcal{C}_{\mathcal{L}} \subset \mathbb{P}^{d-1}$:

➡ What is its defining polynomial $H_{\mathcal{L}}$?

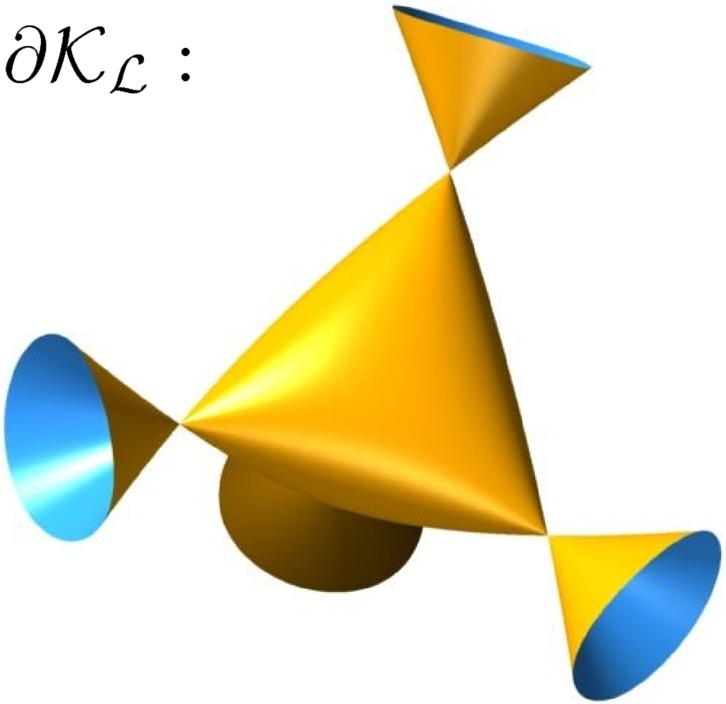
$\partial\mathcal{C}_{\mathcal{L}}$ and its defining polynomial $H_{\mathcal{L}}$

Example:

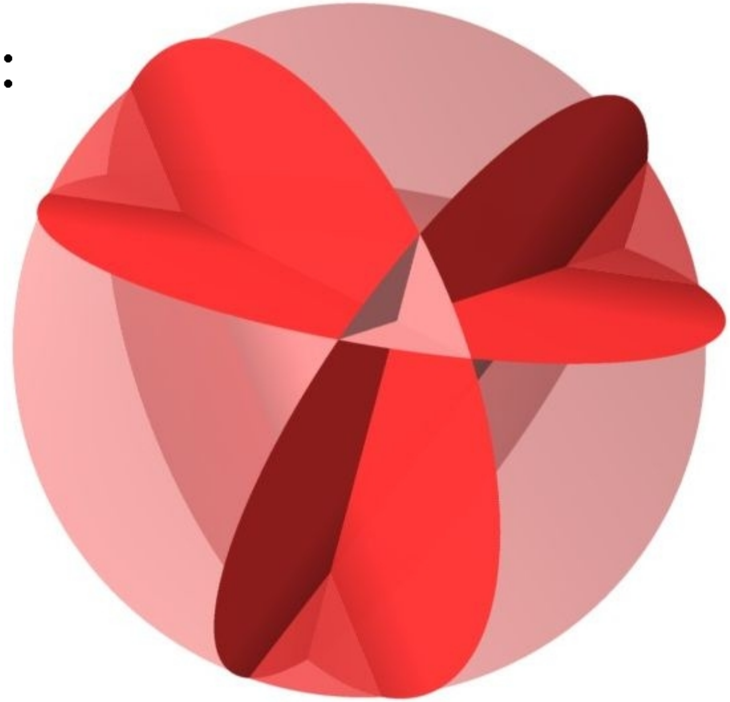
$\mathcal{L}$ given by

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

$\partial\mathcal{K}_{\mathcal{L}}$:

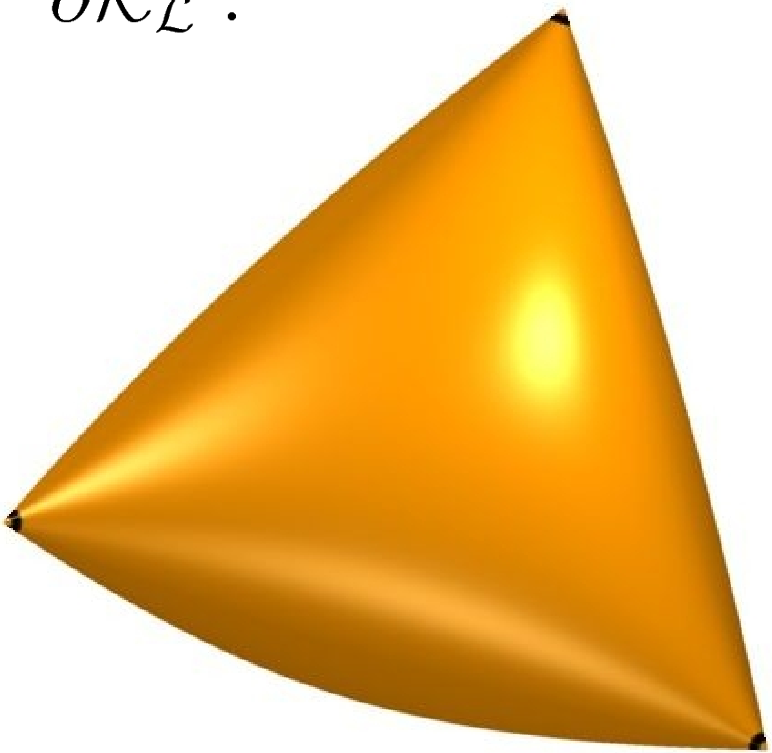


$\partial\mathcal{C}_{\mathcal{L}}$:

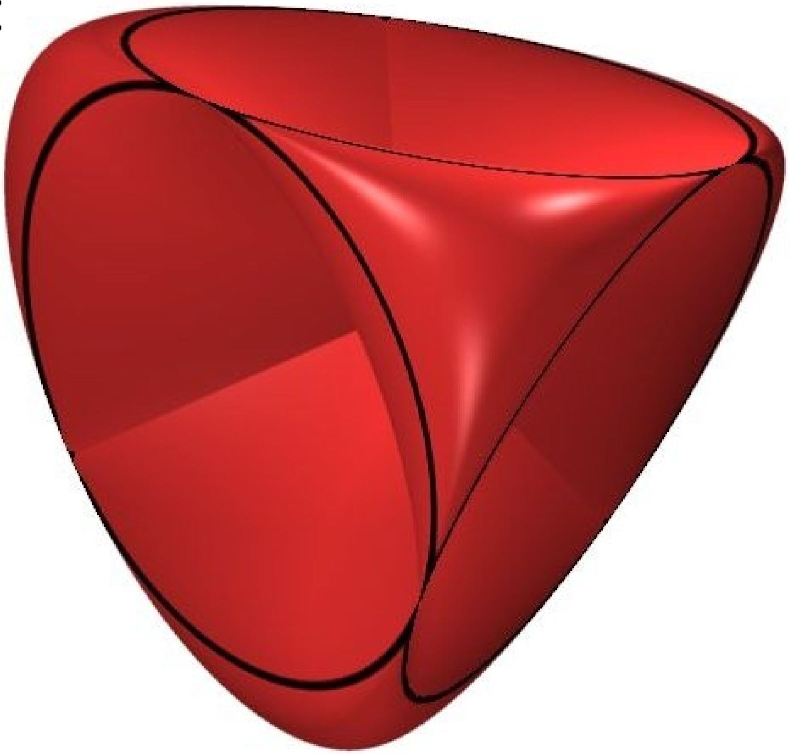


Representing $\partial\mathcal{C}_{\mathcal{L}}$

$\partial\mathcal{K}_{\mathcal{L}}$:



$\partial\mathcal{C}_{\mathcal{L}}$:



- **surfex:** a software for visualizing algebraic surfaces

Magic dualization

Theorem:

Each irreducible hypersurface in the Zariski closure of $\partial\mathcal{C}_{\mathcal{L}}$ is the projectively dual variety to some irreducible component of the hypersurface $\{\det(K)=0\} \subset \mathbb{P}^{d-1}$, or it is dual to some irreducible variety further down in the singularity stratification of $\{\det(K)=0\}$.

Example:

i) Irreducible components: Primary decomposition of ideal of $p \times p$ minors:

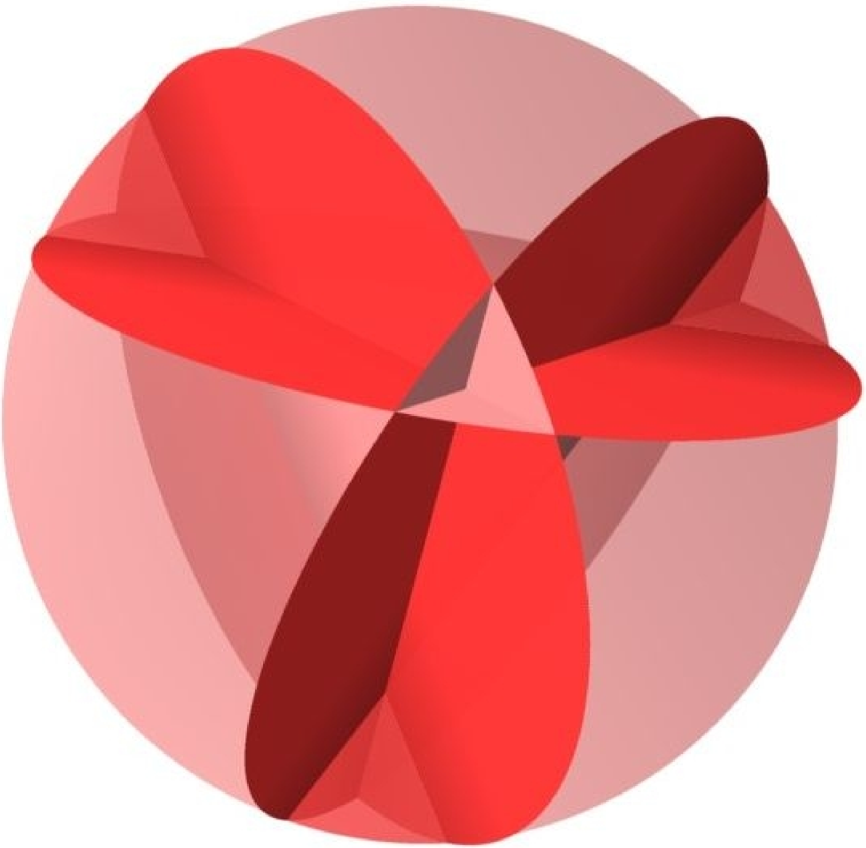
$$(\lambda_3+\lambda_4, \lambda_2+\lambda_4, \lambda_1-\lambda_4),\ \ (\lambda_3+\lambda_4, \lambda_2-\lambda_4, \lambda_1+\lambda_4),\ \ (\lambda_3-\lambda_4, \lambda_2+\lambda_4, \lambda_1+\lambda_4),$$
$$(\lambda_3-\lambda_4, \lambda_2-\lambda_4, \lambda_1-\lambda_4),\ \ (\lambda_1^3-\lambda_1\lambda_2^2-\lambda_1\lambda_3^2+2\lambda_2\lambda_3\lambda_4-\lambda_1\lambda_4^2).$$

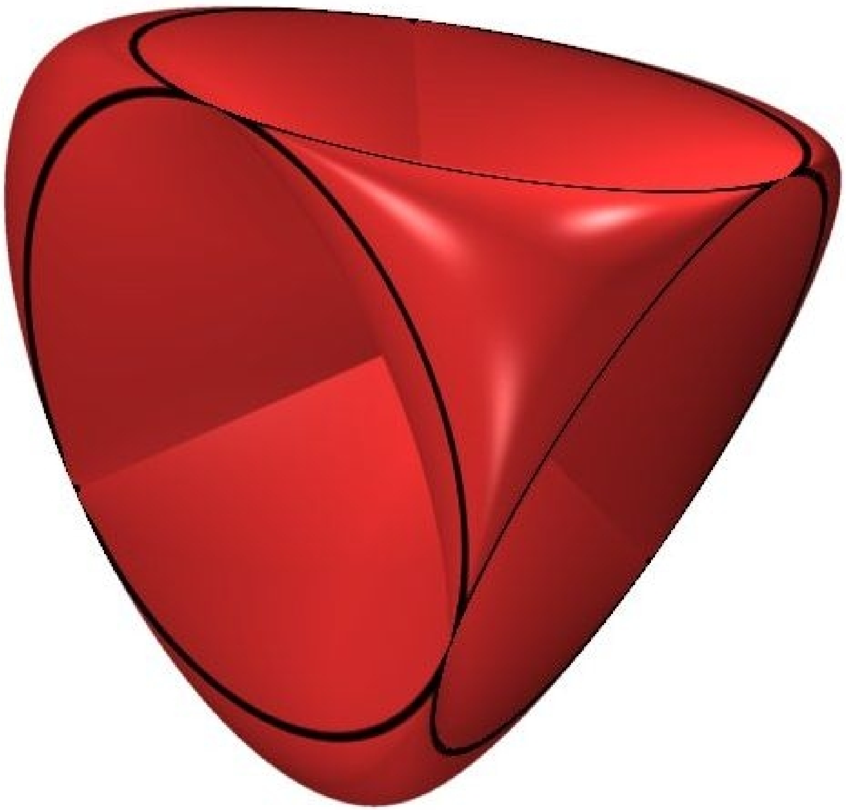
ii) Check which components meet $\partial\mathcal{K}_{\mathcal{L}}$: all 5 components.

iii) Dualize components:

$$H_{\mathcal{L}} = (t_1-t_2-t_3+t_4)\cdot(t_1-t_2+t_3-t_4)\cdot(t_1+t_2-t_3-t_4)\cdot(t_1+t_2+t_3+t_4)$$
$$\cdot(t_2^2t_3^2-2t_1t_2t_3t_4+t_2^2t_4^2+t_3^2t_4^2)$$

Representing $\partial\mathcal{C}_\mathcal{L}$





$$H_\mathcal{L} = (t_1-t_2-t_3+t_4)\cdot(t_1-t_2+t_3-t_4)\cdot(t_1+t_2-t_3-t_4)\cdot(t_1+t_2+t_3+t_4)$$
$$\cdot(t_2^2t_3^2-2t_1t_2t_3t_4+t_2^2t_4^2+t_3^2t_4^2)$$

Gaussian graphical models

- $G = ([m], E)$ undirected graph with $(j,j) \in E \quad \forall j \in [m]$.

- $\mathcal{L} \subset \mathbb{S}^m$ is defined by the linear equations

$$k_{ij} = 0 \quad \text{for} \quad (i,j) \notin E.$$

- $\mathcal{K}_\mathcal{L} = \{K \in \mathbb{S}^m_{\succ 0} \;:\; K_{ij} = 0 \text{ for } (i,j) \notin E\}$
- $\mathcal{C}_\mathcal{L} = \{S_G \in \mathbb{R}^E \;:\; S_G \text{ extendable to pd matrix}\}$
- **ML estimation in Gaussian graphical models = classical pd matrix completion problem**

Colored Gaussian graphical models

- $G = ([m], E)$ undirected graph.
- $V = V_1 \sqcup \cdots \sqcup V_P, \; P \leq m,$ partitioning in vertex color classes.
- $E = E_1 \sqcup \cdots \sqcup E_Q, \; Q \leq |E|,$ partitioning in edge color classes.
- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ colored graph

- $\mathcal{L} \subset \mathbb{S}^m$ is defined by the following linear equations:
  - $k_{ij} = 0$ for $(i,j) \notin E$
  - $k_{ii} = k_{jj}$ for i, j in common vertex color class
  - $k_{ij} = k_{st}$ for $(i,j), (s,t)$ in common edge color class

Colored Gaussian graphical models

- $$\mathcal{K}_{\mathcal{L}} = \left\{ K \in \mathbb{S}^m_{\succ 0} \;:\; \begin{array}{ll} K_{ij} = 0 & \text{for } (i,j) \notin E \\ K_{ii} = K_{jj} & \text{for } i,j \in V_p \\ K_{ij} = K_{st} & \text{for } (i,j),(s,t) \in E_q \end{array} \right\}$$

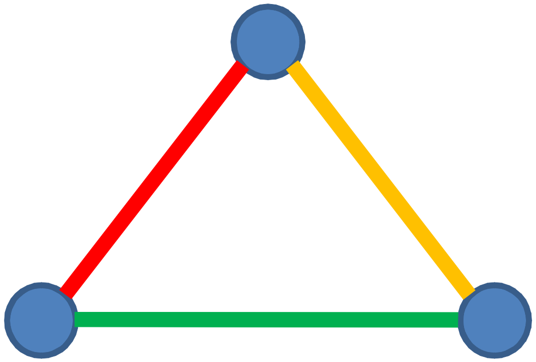
- $$\mathcal{C}_{\mathcal{L}} = \left\{ \left(\sum\nolimits_{i \in V_p} s_{ii}\right)_{p=1,\dots,P}, \left(\sum\nolimits_{(i,j)\in E_q} s_{ij}\right)_{q=1,\dots,Q} \;:\; S \in \mathbb{S}^m_{\succ 0} \right\}$$

- **ML estimation in colored Gaussian graphical models = special pd matrix completion problem:**

  Given $\delta_1, \dots, \delta_P \in \mathbb{R}$ and $\eta_1, \dots, \eta_Q \in \mathbb{R}$, determine whether there exists a PD matrix $\Sigma = (\sigma_{ij})$ with

$$\sum_{i \in V_p} \sigma_{ii} = \delta_p, \quad \forall\, p \qquad \text{and} \qquad \sum_{(i,j)\in E_q} \sigma_{ij} = \eta_q, \quad \forall\, q.$$

Example

$\mathcal{G} =$ 

$$K_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad K_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad K_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

- Our guiding example!

Paper just appeared on the arXiv:

B. Sturmfels, C. U.: Multivariate Gaussians, Semidefinite Matrix Completion, and Convex Algebraic Geometry

arXiv:0906.3529

Thank you!