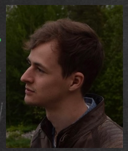
The Maximum Likelihood Degree of Linear Spaces of Symmetric Matrices

Kathlén Kohn
KTH


Carlos Améndola (TU Munich / Uni Ulm)


Lukas Gustafsson (KTH)

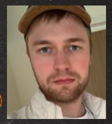

Orlando Marigliano (KTH)


Anna Seigal (Oxford)


Yuhan Jiang (Harvard)


Rosa Winter (MPI-MiS Leipzig)


Stefan Dye (Babbel, Berlin)


Felix Rydell (KTH)


Rainer Sinn (Uni Leipzig)

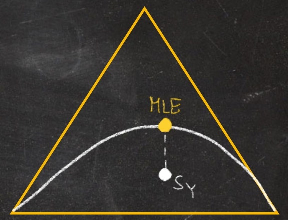
Maximum Likelihood Estimation

Given:
- $\mathcal{M}$: a statistical model = a set of probability distributions
- $Y=(Y_1,\dots,Y_n)$: n samples of observed data

Goal: Find a distribution in the model $\mathcal{M}$ that best fits the empirical data Y.

Approach: maximize the likelihood function

$$L_Y(p) := p(Y_1)\cdots p(Y_n) \quad \text{where } p\in\mathcal{M}.$$



A maximum likelihood estimate (MLE) is a distribution in the model $\mathcal{M}$ that maximizes the likelihood.

Gaussian Models

Density function of m-dimensional Gaussian with mean zero:

$$p_\Sigma(Y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}}\exp\left(-\tfrac{1}{2}Y^T\Sigma^{-1}Y\right) \quad \text{where } Y\in\mathbb{R}^m.$$

Σ = covariance matrix
symmetric, positive definite matrix in $\mathbb{R}^{m\times m}$

$K=\Sigma^{-1}$ = concentration matrix
symmetric, positive definite matrix in $\mathbb{R}^{m\times m}$

A Gaussian model $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m\times m$ symmetric positive definite matrices.

Given data $Y=(Y_1,\dots,Y_n)$, the log-likelihood is

$$\ell_{S_Y}(K) = \log\det(K) - \mathrm{tr}(KS_Y)$$

where $S_Y = \frac{1}{n}\sum_{i=1}^{n} Y_iY_i^T$ is the sample covariance matrix.

Linear Concentration Models

trace inner product: $\langle A,B\rangle = \mathrm{tr}(AB)$

Consider a linear Gaussian model $\mathcal{M}$.
$\Rightarrow$ its Zariski closure in $\mathbb{S}^m = \{K\in\mathbb{C}^{m\times m} \mid K \text{ symmetric}\}$ is a linear subspace $\mathcal{L}\subseteq\mathbb{S}^m$.

The maximum likelihood degree (ML degree) of a linear subspace $\mathcal{L}\subseteq\mathbb{S}^m$ is the number of complex $\underbrace{\text{critical points of } \ell_S|_{\mathcal{L}}}_{K\in\mathcal{L}:\ K^{-1}-S\in\mathcal{L}^\perp}$ for generic $S\in\mathbb{S}^m$.

$\ell_S(K) = \log\det(K) - \mathrm{tr}(KS)$
$\Rightarrow \nabla_K \ell_S = K^{-1}-S$

Assume: $\mathcal{L}$ is regular, i.e. $\exists K\in\mathcal{L}: \det(K)\neq 0$.

Projective Geometry

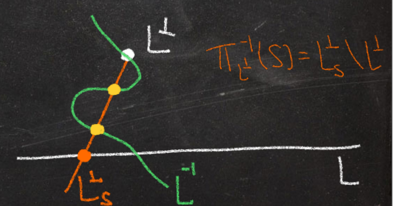
The reciprocal variety $\mathcal{L}^{-1}$ of $\mathcal{L} \subseteq \mathcal{S}^m$ is the Zariski closure of $\{K^{-1} \mid K \in \mathcal{L}\}$ in $\mathcal{S}^m$.

For $\mathcal{L} \subseteq \mathcal{S}^m$, consider projectivizations in $\mathbb{P}\mathcal{S}^m$:

$L := \mathbb{P}\mathcal{L}$,
$L^{-1} := \mathbb{P}\mathcal{L}^{-1}$,
$L^\perp := \mathbb{P}\mathcal{L}^\perp$,
$L_S^\perp := \mathbb{P}\,\mathrm{span}\{\mathcal{L}^\perp, S\}$.

Prop: The ML degree of a linear space $\mathcal{L} \subseteq \mathcal{S}^m$ is $|(L^{-1} \cap L_S^\perp) \setminus L^\perp|$ for generic $S \in \mathbb{P}\mathcal{S}^m$.

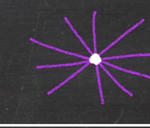
In other words, the ML degree of $\mathcal{L}$ is the degree of $\pi_{L^\perp}|_{L^{-1}}$ where $\pi_{L^\perp} : \mathbb{P}\mathcal{S}^m \dashrightarrow L$ is the projection away from $L^\perp$.



Line Geometry

Consider:

- $\gamma : \overline{\{(\Sigma, \Gamma, \ell) \in L^{-1} \times L^\perp \times Gr(1, \mathbb{P}\mathcal{S}^m) \mid \Sigma \neq \Gamma,\ \Sigma \in \ell,\ \Gamma \in \ell\}} \longrightarrow \mathcal{J} \subseteq Gr(1, \mathbb{P}\mathcal{S}^m)$, $(\Sigma, \Gamma, \ell) \longmapsto \ell$; $\mathcal{J}$ = $\mathrm{im}(\gamma)$
- $\mathcal{G}_S := \{\ell \in Gr(1, \mathbb{P}\mathcal{S}^m) \mid S \in \ell\}$



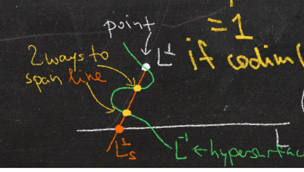
Thm [AGKMS]: The ML degree of a linear space $\mathcal{L} \subseteq \mathcal{S}^m$ is $|\mathcal{J} \cap \mathcal{G}_S| \cdot \deg(\gamma)$ for generic $S \in \mathbb{P}\mathcal{S}^m$.

$\deg(\gamma)$: $= 1$ if $\mathrm{codim}(\mathcal{L}) > 1$

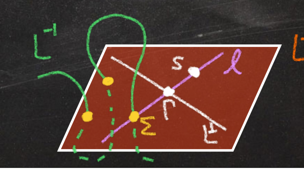
Otherwise:

$|\mathcal{J} \cap \mathcal{G}_S|$: $= 1$ if $\mathrm{codim}(\mathcal{L}) = 1$

$(\mathcal{J} \cap \mathcal{G}_S = \{L_S^\perp\})$



ML degree$(\mathcal{L}) = 0$ $\iff$ $\mathcal{J} \cap \mathcal{G}_S = \emptyset$ for generic $S \in \mathbb{P}\mathcal{S}^m$.



Intersection Theory

Recall: MLdegree$(\mathcal{L}) = |(L^{-1} \cap L_S^\perp) \setminus L^\perp|$ for generic $S \in \mathbb{P}\mathcal{S}^m$

Cor: a) MLdegree$(\mathcal{L}) \leq \deg L^{-1}$

b) If $L^{-1} \cap L^\perp$ is finite & consists only of smooth points of L^{-1}, then MLdegree$(\mathcal{L}) = \deg(L^{-1}) - \deg(L^{-1} \cap L^\perp)$.

↳ How to generalize this?

Intersection Theory

- $L^{-1} \cap L^\perp \cong \Delta \cap (L^{-1} \times L^\perp)$ where Δ is the diagonal in $\mathbb{P}\mathcal{S}^m \times \mathbb{P}\mathcal{S}^m$.
- For $j \in \{0, \dots, \dim(L^{-1} \cap L^\perp)\}$, consider the j-th Segre class (Beniamino Segre (1903–1977))
  $s^j(\Delta \cap (L^{-1} \times L^\perp), L^{-1} \times L^\perp) \in CH_j(\Delta \cap (L^{-1} \times L^\perp))$.
  CH_j: Chow group of j-dimensional cycles
- $\sigma^j(\Delta \cap (L^{-1} \times L^\perp), L^{-1} \times L^\perp)$ = degree of $s^j(\ldots,\ldots)$ under the inclusion $\Delta \cap (L^{-1} \times L^\perp) \hookrightarrow \Delta \cong \mathbb{P}\mathcal{S}^m$.

Thm [AGKMS]: The ML degree of a linear space $\mathcal{L} \subseteq \mathcal{S}^m$ is

$$\deg L^{-1} - \sum_{j=0}^{\delta} \binom{N}{j} \sigma^j(\Delta \cap (L^{-1} \times L^\perp), L^{-1} \times L^\perp)$$

where $\delta := \dim(L^{-1} \cap L^\perp)$ & $N := \dim \mathbb{P}\mathcal{S}^m$.

Extreme Dimensions

① L = point
$\Rightarrow \mathrm{MLdegree}(\mathcal{L}) = 1$

③ L = hyperplane

Thm [AGKMS]: Let $\mathcal{L} = \{K \in \mathbb{S}^m \mid \mathrm{tr}(AK) = 0\}$. ($A \in \mathbb{S}^m \setminus \{0\}$)
$\Rightarrow \mathrm{MLdegree}(\mathcal{L}) = \mathrm{rk}(A) - 1$.

② L = line
$GL(m) \circlearrowright \mathbb{S}^m$ congruence action $\rightsquigarrow GL(m) \circlearrowright Gr(1, \mathbb{P}\mathbb{S}^m)$

- Weierstraß/Segre: geometric classification of $GL(m)$-orbits of lines by Segre symbols
  Corrado Segre (1863–1924)
- Fevola/Mandelshtam/Sturmfels: formula for ML degree in terms of Segre symbols

3x3 matrices

$\begin{bmatrix} a & x & y \\ x & b & z \\ y & z & c \end{bmatrix}$ $\dim \mathbb{P}\mathbb{S}^3 = 5$

① L = point $\Rightarrow \mathrm{MLdegree}(\mathcal{L}) = 1$

② L = line

	[1 1 1]	[2 1]	[(1 1) 1]	[3]	[(2 1)]
$\deg L^{-1}$	2	2	1	2	1
$\mathrm{mld}(\mathcal{L})$	2	1	1	0	0

Segre symbols: 5 congruence classes of regular lines

③ L = 2-plane
∞ congruence classes
1 congruence class each
13 geometric types by CTC Wall

	A	B	B^*	C	D	D^*	E	E^*	F	F^*	G	G^*	H
$\deg L^{-1}$	4	3	4	3	2	4	1	4	2	2	1	2	1
$\mathrm{mld}(\mathcal{L})$	4	3	3	2	2	2	1	1	0	1	0	0	0

DKRS

④ L = 3-plane
singular $\mathcal{L}^\perp$

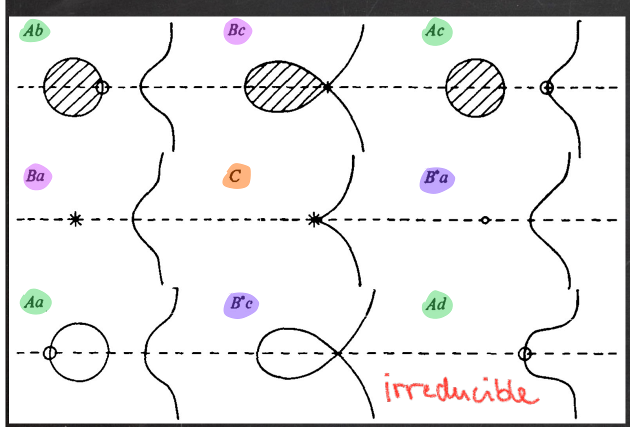
	[111]	[21]	[(11)1]	[3]	[(21)]	[;1;]	[11;;1]	[2;;1]
$\deg L^{-1}$	4	4	4	4	4	1	4	1
$\mathrm{mld}(\mathcal{L})$	4	3	2	2	1	1	1	0

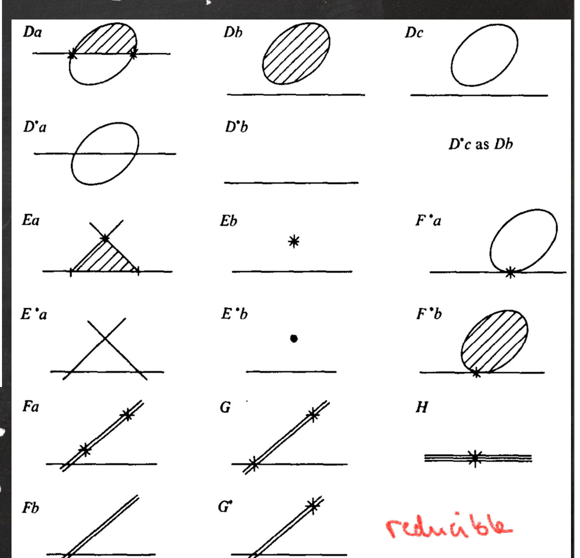
Segre symbol of $\mathcal{L}^\perp$: 8 congruence classes

⑤ L = hyperplane $\Rightarrow \mathrm{MLdegree}(\{K \in \mathbb{S}^3 \mid \mathrm{tr}(AK) = 0\}) = \mathrm{rk}(A) - 1 \in \{0, 1, 2\}$

CTC Wall: Nets of Conics

L = 2-plane in $\mathbb{P}\mathbb{S}^3$





A ≡ smooth
B ≡ node at rank-1 matrix
(Chow hypersurface of rank-1 matrices)
B* ≡ node at rank-2 matrix
(2nd coisotropic hypersurface of Z(det))

$L \cap Z(\det)$
C ≡ cusp
("B ∩ B*")

ML estimation for nets of conics [DKRS]

L^{-1} ≡ Veronese surface in $\mathbb{P}^5$

L^{-1} ≡ projection of Veronese surface from plane ≡ plane

	A	B	B^*	C	D	D^*	E	E^*	F	F^*	G	G^*	H
$\deg L^{-1}$	4	3	4	3	2	4	1	4	2	2	1	2	1
$\mathrm{mld}(\mathcal{L})$	4	3	3	2	2	2	1	1	0	1	0	0	0
$L^{-1} \cap L^\perp$	∅	∅	1 pt.	1pt.	∅	2pts.	∅	3pts.	double pt.	1pt. (smooth)	1pt.	double pt. (smooth)	line

L^{-1} ≡ projection of Veronese surface from point on it ≡ rational normal scroll in $\mathbb{P}^4$

L^{-1} ≡ projection of Veronese surface from secant line ≡ $\mathbb{P}^1 \times \mathbb{P}^1 \subseteq \mathbb{P}^3$

L^{-1} ≡ projection of Veronese surface from tangent line ≡ cone over conic in $\mathbb{P}^3$

Only singular L^{-1}

Extreme ML Degrees

$$0 \le \text{MLdegree}(\mathcal{L}) \le \deg(\overline{\mathcal{L}^{-1}})$$

- $\text{MLdegree}(\mathcal{L}) = \deg(\overline{\mathcal{L}^{-1}}) \iff \overline{\mathcal{L}^{-1}} \cap \mathcal{L}^{\perp} = \emptyset$
- The ML degree of a generic $\mathcal{L} \in Gr(d, \mathbb{S}^m)$ is $\deg(\overline{\mathcal{L}^{-1}})$.
- $NM_{d,m} := \{\mathcal{L} \mid \text{MLdegree}(\mathcal{L}) < \deg(\overline{\mathcal{L}^{-1}})\} \subseteq Gr(d, \mathbb{S}^m)$

  non-maximal

$$\boxed{\text{Thm [JKW]: } NM_{d,m} = Bad_{d,m}}$$

Extreme ML Degrees

$$0 \le \text{MLdegree}(\mathcal{L}) \le \deg(\overline{\mathcal{L}^{-1}})$$

$$\boxed{\text{Thm [JKW]: } NM_{d,m} = Bad_{d,m}}$$

- For $\mathcal{L} \in Gr(d, \mathbb{S}^m_{\mathbb{R}})$, consider projection $\rho : \mathbb{S}^m_{\mathbb{R}} \longrightarrow \text{Hom}(\mathcal{L}, \mathbb{R})$ that is dual to $\mathcal{L} \hookrightarrow \mathbb{S}^m_{\mathbb{R}}$.

  notion due to Gábor Pataki

- $\mathcal{L}$ is called bad if the image of the cone of positive semidefinite matrices under ρ is not closed,

  equivalently, strong duality in semidefinite programming fails for $\mathcal{L}$.
- $Bad_{d,m} := \overline{\{\mathcal{L} \in Gr(d, \mathbb{S}^m_{\mathbb{R}}) \mid \mathcal{L} \text{ is bad}\}} \subseteq Gr(d, \mathbb{S}^m)$

Extreme ML Degrees

$$0 \le \text{MLdegree}(\mathcal{L}) \le \deg(\overline{\mathcal{L}^{-1}})$$

$\overline{\{L \in Gr(d-1, \mathbb{P}\mathbb{S}^m) \mid \exists A \in Reg(\mathbb{P}D_s) \cap L : L + T_A \mathbb{P}D_s \neq \mathbb{P}\mathbb{S}^m\}}$

Cor: $NM_{d,m}$ = union of the coisotropic hypersurfaces in $Gr(d, \mathbb{S}^m)$ associated to the determinantal varieties $D_s := \{A \in \mathbb{S}^m \mid rk(A) \le s\}$ where s ranges over the integers such that $\binom{m-s+1}{2} < d \le \underbrace{\binom{m+1}{2} - \binom{s+1}{2}}_{}$.

proven for $Bad_{d,m}$ by Jiang & Sturmfels

Chow hypersurface

range [incl. $\binom{m-s+1}{2}$] where coisotropic varieties are hypersurfaces

Extreme ML Degrees

$$0 \le \text{MLdegree}(\mathcal{L}) \le \deg(\overline{\mathcal{L}^{-1}})$$

Thm [AGKMS]: The following are equivalent:

a) $\text{MLdegree}(\mathcal{L}) = 0$.

b) $\pi_{L^\perp}|_{\overline{\mathcal{L}^{-1}}} : \overline{\mathcal{L}^{-1}} \dashrightarrow L$ is not dominant.

c) $\text{join}(\overline{\mathcal{L}^{-1}}, L^{\perp}) \neq \mathbb{P}\mathbb{S}^m$.

d) $(K L^{\perp} K) \cap L \neq \emptyset$ for generic $K \in \mathcal{L}$.

e) For bases $\{A_1, \dots, A_c\}$ of $\mathcal{L}^{\perp}$ & $\{B_1, \dots, B_d\}$ of $\mathcal{L}$, $\det(M) \in \mathbb{C}[s_1, \dots, s_d]$ is zero, where $M_{ij} = \sum_{k,l=1}^{d} s_k s_l \cdot tr(A_i B_k A_j B_l)$.

3x3 matrices

$\begin{bmatrix} a & x & y \\ x & b & z \\ y & z & c \end{bmatrix}$ $\dim \mathbb{P}\mathbb{S}^3 = 5$

① L = point $\Rightarrow$ MLdegree$(\mathcal{L}) = 1$

② L = line

	[1 1 1]	[2 1]	[(1 1) 1]	[3]	[(2 1)]
$\deg L^{-1}$	2	2	1	2	1
$\mathrm{mld}(\mathcal{L})$	2	1	1	0	0

L^{-1} & $L^{\perp}$ are contained in a common hyperplane

③ L = 2-plane

	A	B	B^*	C	D	D^*	E	E^*	F	F^*	G	G^*	H
$\deg L^{-1}$	4	3	4	3	2	4	1	4	2	2	1	2	1
$\mathrm{mld}(\mathcal{L})$	4	3	3	2	2	2	1	1	0	1	0	0	0

④ L = 3-plane

	[111]	[21]	[(11)1]	[3]	[(21)]	[;1;]	[11;;1]	[2;;1]
$\deg L^{-1}$	4	4	4	4	4	1	4	1
$\mathrm{mld}(\mathcal{L})$	4	3	2	2	1	1	1	0

here they are not!

⑤ L = hyperplane $\Rightarrow$ MLdegree$(\{K \in \mathbb{S}^3 \mid \mathrm{tr}(AK) = 0\}) = \mathrm{rk}(A) - 1 \in \{0, 1, 2\}$

Final Example

$\mathcal{L} = \{K \in \mathbb{S}^3 \mid \mathrm{tr}(AK) = 0\}$, $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $L^{\perp}$

$\Rightarrow L^{-1} = \mathbb{P}\left\{ \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \;\middle|\; \sigma_{22}\sigma_{33} - \sigma_{23}^2 = 0 \right\}$ quadric cone whose vertex set $(\cong \mathbb{P}^2)$ contains the point $L^{\perp}$

$\Rightarrow \mathrm{join}(L^{-1}, L^{\perp}) = L^{-1}$

but L^{-1} & $L^{\perp}$ are not contained in a common hyperplane