

The Maximum Likelihood Degree of Linear Spaces of Symmetric Matrices

Kathlen Kohn
KTH



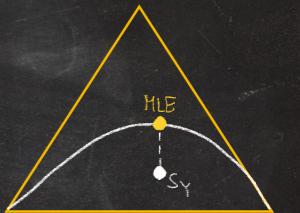
Maximum likelihood Estimation

- Given:
- 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 M that best fits the empirical data Y .

Approach: maximize the likelihood function

$$L_p(Y) := p(Y_1) \cdots p(Y_n) \text{ where } p \in M.$$



A maximum likelihood estimate (MLE) is a distribution in the model M that maximizes the likelihood.

Gaussian Models

Density function of m -dimensional Gaussian with mean zero:

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

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 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

$$l_S(K) = \log \det(K) - \text{tr}(KS)$$

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

Linear Concentration Models

Consider a linear Gaussian model M .

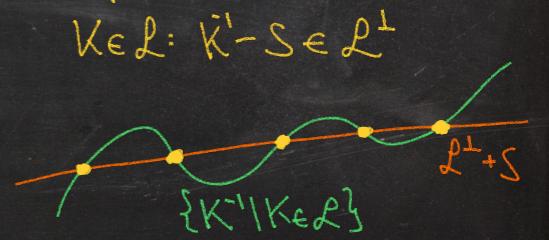
\Rightarrow its Zariski closure in $\mathbb{S}^m = \{K \in \mathbb{C}^{m \times m} \mid K \text{ symmetric}\}$
is a linear subspace $L \subseteq \mathbb{S}^m$.

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

The maximum likelihood degree (ML degree) of a linear subspace $L \subseteq \mathbb{S}^m$
is the number of complex critical points of $l_S|_L$ for generic $S \in \mathbb{S}^m$.

$$\begin{aligned} l_S(K) &= \log \det(K) - \text{tr}(KS) \\ \Rightarrow \nabla_K l_S &= K^{-1} - S \end{aligned}$$

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



Projective Geometry

The reciprocal variety \tilde{L}^{\perp} of $L \subseteq \mathbb{P}^m$ is the Zariski closure of $\{\tilde{K}' \mid K \in L\}$ in \mathbb{P}^m .



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

$$L' := \mathbb{P}L^{\perp},$$

$$L^{\perp} := \mathbb{P}L^{\perp},$$

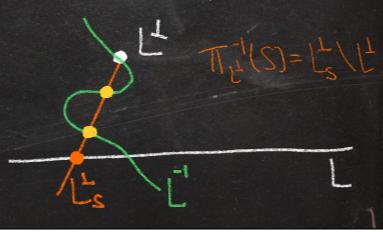
$$L_s^{\perp} := \mathbb{P} \text{span}\{L^{\perp}, S\}.$$

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

In other words, the ML degree of L is the degree of $\pi_{L'}^{-1}|_{L_s^{\perp}}$ where

$$\pi_{L'}: \mathbb{P}\mathbb{P}^m \dashrightarrow L'$$

is the projection away from L^{\perp} .



Line Geometry

Consider:

$$\bullet \gamma: \overline{\{(\Sigma, \Gamma, \ell) \in L' \times L^{\perp} \times \text{Gr}(1, \mathbb{P}\mathbb{P}^m) \mid \Sigma \neq \Gamma, \Sigma \in \ell, \Gamma \in \ell\}} \xrightarrow{\quad j \in \text{Gr}(1, \mathbb{P}\mathbb{P}^m)} (\Sigma, \Gamma, \ell) \mapsto \ell$$

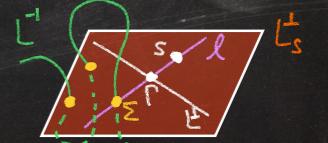
$$\bullet G_S := \{ \ell \in \text{Gr}(1, \mathbb{P}\mathbb{P}^m) \mid S \in \ell \}$$

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

Otherwise:

ML degree(L) = 0
 $\Leftrightarrow j \cap G_S = \emptyset$
 for generic $S \in \mathbb{P}\mathbb{P}^m$

$= 1$ if $\text{codim}(L) > 1$
 point
 2 ways to span line
 ℓ_s^{\perp} if $\text{codim}(L) = 1$



Intersection Theory



Recall: MLdegree(L) = $|(L' \cap L_s^{\perp}) \setminus L'|$ for generic $S \in \mathbb{P}\mathbb{P}^m$

(Cor: a) MLdegree(L) $\leq \deg(L')$

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

↪ How to generalize this?

Intersection Theory

Benedetto Scare
(1903-1977)

- $L' \cap L_s^{\perp} \cong \Delta \cap (L' \times L_s^{\perp})$ where Δ is the diagonal in $\mathbb{P}\mathbb{P}^m \times \mathbb{P}\mathbb{P}^m$.
- For $j \in \{0, \dots, \dim(L' \cap L_s^{\perp})\}$, consider the j -th Segre class $s_j^{\delta}(\Delta \cap (L' \times L_s^{\perp}), L' \times L_s^{\perp}) \in CH_j(\Delta \cap (L' \times L_s^{\perp}))$.
Chow group of j -dimensional cycles
- $\sigma^{\delta}(\Delta \cap (L' \times L_s^{\perp}), L' \times L_s^{\perp}) = \text{degree of } s_j^{\delta}(\Delta \cap (L' \times L_s^{\perp}), L' \times L_s^{\perp})$ under the inclusion $\Delta \cap (L' \times L_s^{\perp}) \hookrightarrow \Delta \cong \mathbb{P}\mathbb{P}^m$.

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

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

where $S := \dim(L' \cap L_s^{\perp})$ & $N := \dim \mathbb{P}\mathbb{P}^m$.

Extreme Dimensions

① $L = \text{point}$

$$\Rightarrow \text{MLdegree}(L) = 1$$

③ $L = \text{hyperplane}$

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

② $L = \text{line}$

$GL(m) \curvearrowright \mathbb{S}^m$ congruence action $\rightsquigarrow GL(m) \curvearrowright \text{Gr}(1, \mathbb{P}\mathbb{S}^m)$

- Weierstraß/Segre: geometric classification of $GL(m)$ -orbits of lines by Segre symbols Corrado Segre (1863-1924)
- Ferrara/Mandelstam/Sturmfels: formula for MLdegree in terms of Segre symbols

1

3x3 matrices

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

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

② $L = \text{line}$

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

Segre symbols:
5 congruence classes
of regular lines

③ $L = 2\text{-plane}$

13 geometric types by CTC Wall

	congruence classes												
	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
$\text{mld}(\mathcal{L})$	4	3	3	2	2	2	1	1	0	1	0	0	0

DKRS

④ $L = 3\text{-plane}$

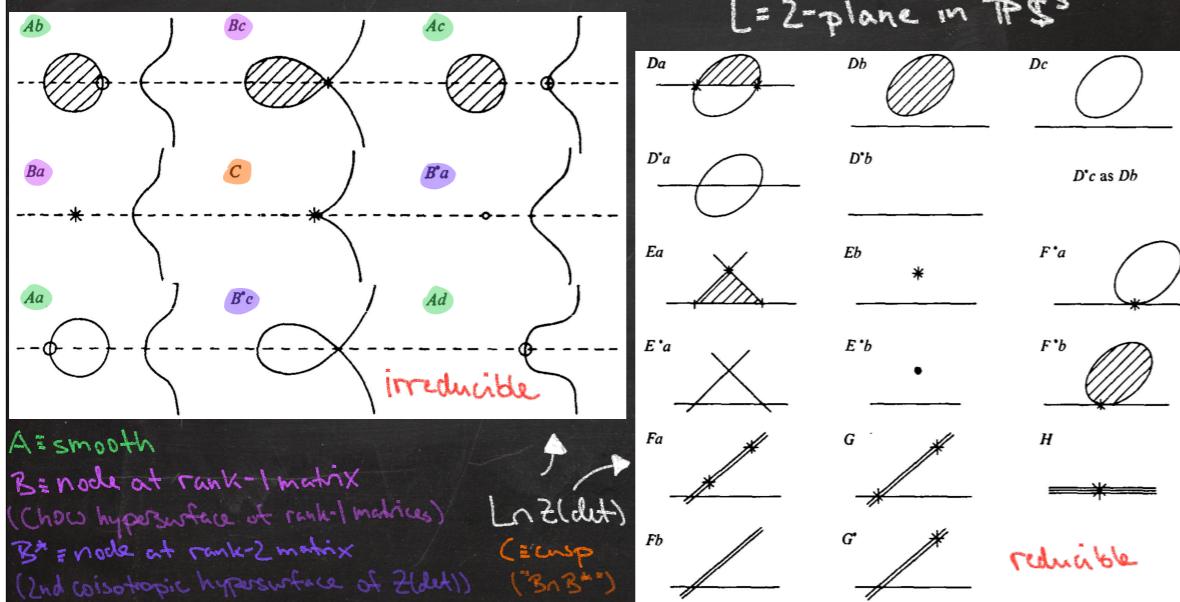
singular L^\perp

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

Segre symbol of L^\perp :
8 congruence classes

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

CTC Wall: Nets of Conics



ML estimation for nets of conics [DKRS]

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

$L^\perp \cap L^\perp = \emptyset$ (smooth line)
 $L^\perp \cap L^\perp = 1 \text{ pt.}$ (smooth point)
 $L^\perp \cap L^\perp = 2 \text{ pts.}$ (smooth double point)
 $L^\perp \cap L^\perp = 3 \text{ pts.}$ (smooth triple point)
 $L^\perp \cap L^\perp = 1 \text{ pt.}$ (smooth double point)
 $L^\perp \cap L^\perp = 1 \text{ pt.}$ (smooth triple point)

$L^\perp = \text{projection of Veronese surface from point on it}$
 $= \text{rational normal scroll in } \mathbb{P}^4$

$L^\perp = \text{projection of Veronese surface from tangent line}$
 $= \text{cone over conic in } \mathbb{P}^3$
 $L^\perp = \text{projection of Veronese surface from secant line}$
 $= \mathbb{P}^1 \times \mathbb{P}^1 \subset \mathbb{P}^3$
 Only singular L^\perp

Extreme ML Degrees

$0 \leq \text{MLdegree}(\mathcal{L}) \leq \deg(\mathcal{L}')$

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

$$\text{Thm [JKW]: } NM_{d,m} = \text{Bad}_{d,m}$$

Extreme ML Degrees

$0 \leq \text{MLdegree}(\mathcal{L}) \leq \deg(\mathcal{L}')$

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

Proven for $\text{Bad}_{d,m}$ by Jiang & Sturmfels

Chow hypersurface

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

Extreme ML Degrees

$0 \leq \text{MLdegree}(\mathcal{L}) \leq \deg(\mathcal{L}')$

$$\text{Thm [JKW]: } NM_{d,m} = \text{Bad}_{d,m}$$

- For $L \in \text{Gr}(d, \mathbb{S}^m_R)$, consider projection $\beta: \mathbb{S}^m_R \rightarrow \text{Hom}(L, \mathbb{R})$
notion due to Gábor Pataki
 that is dual to $L \hookrightarrow \mathbb{S}^m_R$.
- L is called **bad** if the image of the cone of positive semi-definite matrices under β is not closed,
 equivalently, strong duality in semi-definite programming fails for L .
- $\text{Bad}_{d,m} := \overline{\{L \in \text{Gr}(d, \mathbb{S}^m_R) \mid L \text{ is bad}\}} \subseteq \text{Gr}(d, \mathbb{S}^m)$

Extreme ML Degrees

$0 \leq \text{MLdegree}(\mathcal{L}) \leq \deg(\mathcal{L}')$

Thm [AGKMS]: The following are equivalent:

- $\text{MLdegree}(\mathcal{L}) = 0$.
- $\pi|_{\mathcal{L}^\perp}: \mathcal{L}' \dashrightarrow \mathcal{L}$ is not dominant.
- $\text{join}(\mathcal{L}', \mathcal{L}^\perp) \neq \mathbb{P}\mathbb{S}^m$.
- $(KL^\perp K) \cap \mathcal{L} \neq \emptyset$ for generic $K \in \mathcal{L}$.
- For bases $\{A_1, \dots, A_d\}$ 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
 $M_{ij} = \sum_{k,l=1}^d s_k s_l \cdot \text{tr}(A_i B_k A_j B_l)$.

3x3 matrices $\begin{bmatrix} a & x & y \\ x & b & z \\ y & z & c \end{bmatrix}$ $\dim \mathbb{P}\mathbb{S}^3 = 5$

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

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

② $L = \text{line}$

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

L^\perp & L^\perp are contained in a common hyperplane

③ $L = 2\text{-plane}$

	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	0	1
$\text{mld}(\mathcal{L})$	4	3	3	2	2	2	1	1	0	1	0	0	0

④ $L = 3\text{-plane}$

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

here they are not!

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

Final Example

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

$$\Rightarrow L^\perp = \mathbb{P} \left\{ \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \mid \sigma_{22}\sigma_{33} - \sigma_{23}^2 = 0 \right\} \rightsquigarrow \text{quadric cone}$$

whose vertex set ($\cong \mathbb{P}^2$) contains the point L^\perp

$$\Rightarrow \text{join}(L^\perp, L^\perp) = L^\perp$$

but L^\perp & L^\perp are not contained in a common hyperplane