# Maximum likelihood thresholds of linear covariance models

Daniel Irving Bernstein

Tulane University

j.w.w. Shlomo Gortler and Louis Theran:
https://arxiv.org/abs/2305.06280

j.w.w. Sean Dewar, Shlomo Gortler, Tony Nixon, Meera Sitharam, and Louis Theran:
https://arxiv.org/abs/2210.11081
https://arxiv.org/abs/2108.02185

j.w.w. Hayden Outlaw
https://arxiv.org/pdf/2312.03145

# Talk plan

- Definitions
- Reformulation of the MLT of a graphical model as a rigidity-theoretic property of the graph and implications
- What is the "expected" MLT of a Gaussian graphical model?
- Generalize the rigidity-theoretic picture, determine MLTs of "generic" linear concentration models

# Linear concentration models

**tl;dr:**

Linear concentration model = linear space of symmetric matrices that contains a positive definite matrix

- A ***statistical model*** is a family of joint probability distributions
- The $n$-***variate normal distribution*** with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \succ 0$ is the normal distribution with density function

$$f_{\mu,\Sigma}(x) := \frac{\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))}{\sqrt{(2\pi)^v \det(\Sigma)}}$$

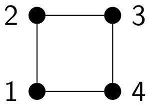
- An $n$-variate ***linear concentration model*** is the family of multivariate normal distributions whose inverse covariance matrix lies in a particular linear subspace of $n \times n$ symmetric matrices
- Covariance matrices are positive definite, so if $L \subseteq \mathcal{S}^n$ does not have a positive definite matrix, the statistical model is empty

# Example

## Definition

Let $G$ be a graph on vertex set $\{1, \ldots, n\}$. Consider the following subspace $L_G \subseteq \mathcal{S}^n$ of $n \times n$ symmetric matrices

$$L_G := \{M \in \mathcal{S}^n : M_{ij} = 0 \text{ whenever } ij \text{ is not an edge of } G\}.$$

$G =$ 


$$\begin{pmatrix} x_{11} & x_{12} & 0 & x_{14} \\ x_{12} & x_{22} & x_{23} & 0 \\ 0 & x_{23} & x_{33} & x_{34} \\ x_{14} & 0 & x_{34} & x_{44} \end{pmatrix}$$

## Definition

The ***Gaussian graphical model*** $\mathcal{M}_G$ associated to a graph $G$ is the linear concentration model associated to $L_G$.

# Maximum likelihood thresholds

Suppose we are given:

- A linear subspace $L \subseteq \mathcal{S}^n$ of $n \times n$ symmetric matrices, and
- a matrix $X \in \mathbb{R}^{n \times d}$, columns supposedly iid from a distribution in $\mathcal{M}_L$

The ***maximum likelihood estimate (MLE)*** is the solution to the following optimization problem, if it exists:

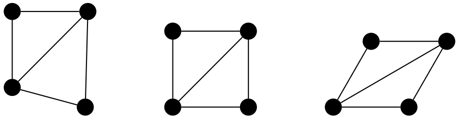
$$\begin{aligned} \max_{K} \quad & \mathrm{Tr}(XX^T K) + \log\det K \\ \text{s.t.} \quad & K \succ 0 \quad \text{and} \quad K \in L. \end{aligned}$$

The ***maximum likelihood threshold of*** $L$ is the minimum $n$ such that the above optimization problem has a solution for any generic $X \in \mathbb{R}^{n \times d}$.

# Rigidity theory basics

## Definition

A ***bar and joint framework in*** $d$ ***dimensions*** consists of a graph $G$, and a map $p : V(G) \to \mathbb{R}^d$. Such a framework is ***independent*** if the edge-lengths can be independently perturbed.



## Theorem (Asimov and Roth 1978)

*Given a graph $G$, if $p : V(G) \to \mathbb{R}^d$ is "generic," then whether the framework $(G, p)$ is independent in $\mathbb{R}^d$ does not depend on $p$.*

One says that $G$ is (generically) independent in $\mathbb{R}^d$ if $(G, p)$ is independent for all generic $p : V(G) \to \mathbb{R}^d$.

# Upper bounds via generic independence

## Definition (Generic completion rank)

The ***generic completion rank of*** $G$, denoted $\mathrm{GCR}(G)$, is the minimum $d$ such that $G$ is generically independent in $\mathbb{R}^{d-1}$.

## Theorem (Uhler 2012, Gross and Sullivant 2018)

$\mathrm{MLT}(G) \leq \mathrm{GCR}(G)$.

$\mathrm{GCR}(G)$ can be computed in RP time, so it would be great if the above inequality were sharp. However...

## Theorem (Blekherman and Sinn 2019)

$\mathrm{MLT}(K_{5,5}) = 4$ *but* $\mathrm{GCR}(K_{5,5}) = 5$.
$\mathrm{MLT}(K_{n,n}) = O(\sqrt{n})$ *whereas* $\mathrm{GCR}(K_{n,n}) = O(n)$.

# MLT in rigidity-theoretic terms

**Definition**

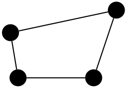
Two frameworks $(G, p)$ and $(G, q)$ on the same graph $G$ are ***equivalent*** if the following holds for every edge $uv$ of $G$

$$\|p(u) - p(v)\| = \|q(u) - q(v)\|.$$

$(G, p)$ in $\mathbb{R}^k$ has ***full affine span*** if $\{p(v) : v \in V(G)\}$ affinely spans $\mathbb{R}^k$.

**Theorem (BDGNST)**

*Let $G$ be a graph with $n$ vertices.* MLT$(G)$ *is the smallest $d$ such that every generic framework in $\mathbb{R}^{d-1}$ is equivalent to a framework in $\mathbb{R}^{n-1}$ with full affine span.*



# More rigidity

## Definition

Two frameworks $(G, p)$ and $(G, q)$ on the same graph $G$ are ***equivalent*** if the following holds for every edge $uv$ of $G$

$$\|p(u) - p(v)\| = \|q(u) - q(v)\|.$$

They are moreover ***congruent*** if the above holds for all pairs $uv$ of vertices of $G$.

## Definition

A framework $(G, p)$ in $\mathbb{R}^d$ is ***globally rigid*** if all equivalent frameworks in $\mathbb{R}^d$ are congruent.

# Lower bounds via global rigidity

**Theorem (Connelly 2005; Gortler, Healy, and Thurston 2010)**

*Given a graph $G$, if $p : V(G) \to \mathbb{R}^d$ is "generic," then whether the framework $(G, p)$ is globally rigid in $\mathbb{R}^d$ does not depend on $p$.*

One says that $G$ is (generically) globally rigid in $\mathbb{R}^d$ if $(G, p)$ is globally rigid for all generic $p : V(G) \to \mathbb{R}^d$.

**Theorem (BDGNST)**

*If a subgraph of $G$ on at least $d + 1$ vertices is generically globally rigid in $\mathbb{R}^{d-1}$, then $\mathrm{MLT}(G) > d$.*

**Theorem (BDGNST)**

$\mathrm{MLT}(G) = \mathrm{GCR}(G)$ *in all of the following cases:*

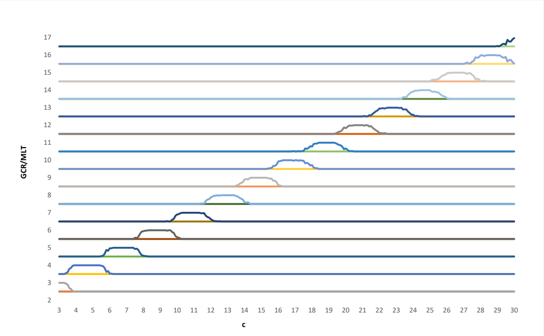
- $\mathrm{GCR}(G) \leq 4$
- $\mathrm{MLT}(G) \leq 3$
- *$G$ has 9 or fewer vertices.*

## The MLT of an Erdös-Renyi random graph

Let $c > 0$ be fixed and let $G(n, n/c)$ denote the Erdös-Renyi random graph model with expected edge density $c$.

### Conjecture (BDGNST)

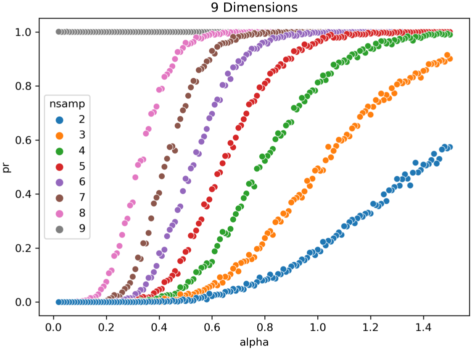
*Let $c > 0$ be fixed. If $G \sim G(n, n/c)$ then $MLT(G) = GCR(G)$ with high probability. In this case, $MLT(G) \approx \lfloor c/2 \rfloor + 2$ with high probability.*



# MLT of GraphLasso outputs

$$\max_{K} \quad \mathrm{Tr}(XX^T K) + \log\det K + \alpha \|K\|_1$$
$$\text{s.t.} \quad K \succ 0.$$

Given $n$ datapoints, i.i.d. from a "reasonable" distribution, with what probability does GraphLasso pick a graphical model with MLT $\leq n$?



# More rigidity: stress matrices

Given a graph $G = (V, E)$ and $p \in (\mathbb{R}^d)^n$, a ***stress*** of the framework $(G, p)$ is a vector $\omega \in \mathbb{R}^E$ such that

$$\sum_{j: ij \in E} \omega_{ij}(p(j) - p(i)) = 0 \qquad \text{for all } i \in V.$$

The associated ***stress matrix*** $\Omega$ is defined by

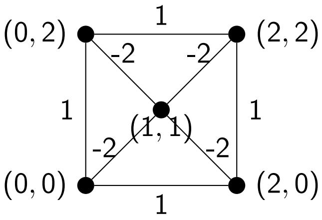
$$\Omega_{ij} = \begin{cases} \omega_{ij} & \text{if } ij \in E \\ 0 & \text{if } i \neq j \text{ and } ij \notin E \\ -\sum_{k: ik \in E} \omega_{ik} & \text{if } i = j. \end{cases}$$

By construction, if $\omega$ is a stress of $(G, p)$ with stress matrix $\Omega$, then

$$\Omega \begin{pmatrix} 1 & -p(1)- \\ 1 & -p(2)- \\ \vdots & \vdots \\ 1 & -p(n)- \end{pmatrix} = 0$$

So $\operatorname{rank}(\Omega) \leq n - d - 1$ if $p$ is generic. This inequality can be strict.

# Stress matrix example




$$\begin{pmatrix} 0 & 1 & 0 & 1 & -2 \\ 1 & 0 & 1 & 0 & -2 \\ 0 & 1 & 0 & 1 & -2 \\ 1 & 0 & 1 & 0 & -2 \\ -2 & -2 & -2 & -2 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 2 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

# Stress matrices in rigidity

**Theorem (Connelly 2005; Gortlear-Healy-Thurston 2010)**

*A graph $G$ on $n$ vertices is generically globally rigid in $\mathbb{R}^d$ iff there exists $p \in (\mathbb{R}^d)^n$ such that $(G, p)$ has a stress matrix of rank $n - d - 1$.*

**Theorem (Alfakih 2011)**

*If $G$ is a graph on $n$ vertices and $p \in (\mathbb{R}^d)^n$, and $(G, p)$ has a PSD stress matrix of rank $n - (d + k) - 1$, then any framework equivalent to $(G, p)$ in $\mathbb{R}^{n-1}$ has an affine span of dimension at most $d + k$.*

**Theorem (BDGNST)**

*Let $G$ be a graph with $n$ vertices.* $\mathrm{MLT}(G)$ *is the smallest $d$ such that every generic framework in $\mathbb{R}^{d-1}$ is equivalent to a framework in $\mathbb{R}^{n-1}$ with full affine span.*

The upshot: the MLT of $G$ is the minimum $d$ such that no generic $d$-dimensional framework $(G, p)$ has a nonzero PSD stress.

# Generalizing to linear concentration models

Let $L \subseteq \mathcal{S}^n$ be a linear subspace of symmetric matrices and let $d \leq n$. Does there exist an open neighborhood $U \subseteq \mathbb{R}^{n \times d}$ such that for each $X \in U$, there exists a ***nonzero PSD*** $\Omega \in L$ such that $\Omega X = 0$?

**Theorem (BGT)**

*The minimum $d$ for which the above answer is "no" is the MLT of $L$.*

Rigidity-theoretic interpretation:

- $L$ is like a graph, $X$ is a point configuration
- projecting $XX^T$ onto $L$ is recording edge-lengths of a framework
- each $\Omega \in L$ satisfying $\Omega X = 0$ is a stress of $X$

When $L = L_G$, we are literally asking if there exists an open neighborhood of $(d-1)$-dimensional frameworks on $G$ that all have a PSD stress.

# A sufficient condition for "no"

Does there exist an open neighborhood $U \subseteq \mathbb{R}^{n\times d}$ such that for each $X \in U$, there exists a ***nonzero PSD*** $\Omega \in L$ such that $\Omega X = 0$?

Let $\mathcal{S}^n(d)$ denote the variety of $n \times n$ symmetric matrices of rank $d$.

## Proposition (BGT)

*If the projection of $\mathcal{S}^n(d)$ onto $L$ is $\dim(L)$-dimensional, the answer is "no."*

- $\dim(L) \leq dn - \binom{d}{2} = \dim(\mathcal{S}^n(d))$ is neccessary for this projection to be full-dimensional.
- When $L = L_G$, this projection is full-dimensional iff $G$ is independent in the $(d-1)$-rigidity matroid.
- Answer for $L = L_G$ can still be "no" even when $G$ is ***not*** independent (Blekherman and Sinn).

# Generic linear spaces

**Theorem (BGT)**

*Let $L \subseteq \mathcal{S}^n$ be a **generic** linear subspace. The following are equivalent:*

1. $\dim(L) \leq nd - \binom{d}{2}$.
2. *The projection of $\mathcal{S}^n(d)$ onto $L$ is $\dim(L)$-dimensional,*
3. *There does **not** exist an open neighborhood $U \subseteq \mathbb{R}^{n \times d}$, and such that for every $X \in U$, there exists a nonzero PSD $\Omega \in L$ such that $\Omega X = 0$.*

**Corollary**

*Let $L \subseteq \mathcal{S}^n$ be a **generic** linear subspace. The MLT of $L$ is the minimum $d$ such that $\dim(L) \leq dn - \binom{d}{2}$.*

- This is the behavior one would expect from a dimension count. Not obvious since we're working over the reals!
- Proof uses tools from convexity theory and differential topology.

# Some remarks on the proof: hard direction

**Lemma**

*Let $L \subseteq \mathcal{S}^n$ be a generic linear subspace. Assume $\dim(L) > nd - \binom{d}{2}$. Then there exists an open neighborhood $U \subseteq \mathbb{R}^{n \times d}$ such that for every $X \in U$, there exists a nonzero PSD $\Omega \in L$ such that $\Omega X = 0$.*

$L$ being "generic" excludes the following two possibilities, neither of which is robust to perturbation

- $L$ does not contain a PSD matrix of rank $n - d$, where $d$ satisfies

$$m \geq dn - \binom{d}{2} - n + d + 1$$

- $L$ contains a PSD matrix of the above rank, but for every such matrix $\Omega$, $L$ non-transversely intersects the space of matrices sharing a kernel with $\Omega$.

# Remarks on the proof: easy direction

**Lemma**

*Let $L \subseteq \mathcal{S}^n$ be a generic linear subspace. Assume $\dim(L) \leq nd - \binom{d}{2}$. Then the projection of $\mathcal{S}^n(d)$ onto $L$ is full-dimensional in $L$.*

This also holds when $L = L_G$ for a graph $G$ that is independent in the $(d-1)$-dimensional rigidity matroid.

Proof idea:

- generic normal spaces of $\mathcal{S}^n(d)$ have dimension $\binom{n-d+1}{2}$
- dimension count says intersection with $L$ expected to be trivial
- if $L$ and normal space have larger than expected dimension, a perturbation of $L$ does not.

# Weak maximum likelihood thresholds

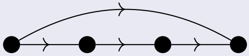
## Definition

The ***weak*** maximum likelihood threshold of a linear concentration model $L \subseteq \mathcal{S}^n$ is the minimum $d$ such that ***some*** open set $U \subseteq \mathbb{R}^{n \times d}$ satisfies the property that for every $X \in U$, the following has a solution

$$\max_{K} \quad \mathrm{Tr}(XX^T K) + \log\det K$$
$$\text{s.t.} \quad K \succ 0 \quad \text{and} \quad K \in L.$$

## Conjecture

*The weak maximum likelihood threshold of a graph $G$ is 2 if and only if it has at least one edge, and an acyclic orientation with no "stretched cycles"*



# The end

Thank you for your attention!

https://arxiv.org/abs/2305.06280
https://arxiv.org/abs/2210.11081
https://arxiv.org/abs/2108.02185
https://arxiv.org/pdf/2312.03145