

Learning the structure of graphical models by covariance queries

Gábor Lugosi

based on joint work with

Jakub Truskowski

Vasiliki Velona

Piotr Zwiernik

gaussian graphical models

Graphical models are often used to represent the dependency structure of multivariate distributions.

We consider **Gaussian** vectors $\mathbf{X} = (X_1, \dots, X_n)$ with covariance matrix Σ .

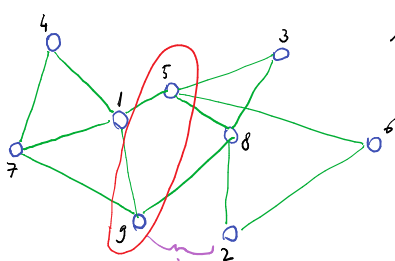
Let $\mathbf{K} = \Sigma^{-1}$ be the **precision matrix**. A basic property of Gaussian vectors is

$$K_{ij} = 0 \iff X_i \perp\!\!\!\perp X_j \mid X_{[n] \setminus \{i,j\}}$$

We define the **concentration graph** $\mathcal{G}(\Sigma) = (V, E)$ with $V = [n]$ as the graph with

$$E = \{i \sim j \mid K_{ij} \neq 0\}$$

gaussian graphical models



$$X = (X_1, \dots, X_n) \sim \mathcal{N}(\mu, \Sigma)$$

$$K = \Sigma^{-1}$$

no edge $\Leftrightarrow K_{2,9} = 0$

$\{9, 5\}$ separates $\{1, 4, 7\}$ and $\{2, 3, 6, 8\}$

(X_1, X_4, X_7) and (X_2, X_3, X_6, X_8) are conditionally independent
given (X_5, X_9)

learning the structure

We address the problem of **structure recovery**: learning the concentration graph from data.

This problem has been extensively studied. Low sample and time complexity algorithms are in the focus.

All algorithms start by computing the **sample covariance matrix**.

Problem: when the dimension n is large, even writing down the covariance matrix is expensive.

Question: is it possible to recover the graph structure with less than quadratic complexity?

Obviously, the question is only meaningful if the graph is sparse.

covariance oracle

Abstract model: We assume access to data through a **covariance oracle**.

The oracle takes a pair $i, j \in [n]$ as an input and outputs Σ_{ij} .

Learn the support of $K = \Sigma^{-1}$ with few oracle queries.

Under what conditions is this possible with $\mathbf{o}(n^2)$ queries?

An easy counterexample: one needs $\Omega(n^2)$ queries to tell the empty graph apart from a graph with a single edge.

We assume that $\mathcal{G}(\Sigma)$ is **connected**.

The simplest connected graphs are **trees**.

trees

When $\mathcal{G}(\Sigma)$ is a tree T , then for every $i, j \in V$, the correlation $\rho_{ij} = \text{corr}(X_i, X_j)$ equals

$$\rho_{ij} = \prod_{uv \in \overline{ij}} \rho_{uv},$$

where $\overline{ij}$ is the unique path between i and j in T .

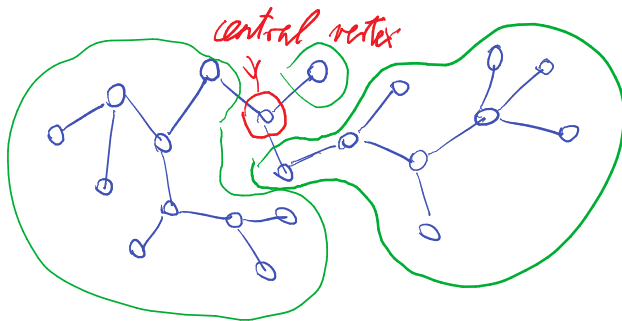
This implies that T is a maximum-weight spanning tree that can be found simply (Chow-Liu, 1968).

However, to run the Chow-Liu algorithm, all correlations must be known—quadratic complexity!

recovering trees: divide and conquer

Randomized algorithm.

- Find a central vertex v .
- Identify the connected components of $T \setminus \{v\}$.
- Recurse.



centrality

The **centroid** of a tree T is a vertex that minimizes

$$c(v) = \max_{C \in \mathcal{C}^v} |C|$$

where $\mathcal{C}^v$ is the set of connected components of $G \setminus v$.

It is unclear how one can find this vertex using covariance queries.

A modified measure of centrality:

$$\tilde{c}(v) = \sum_{C \in \mathcal{C}^v} |C|^2$$

This can be approximated:

- pick k vertices $u_1, \dots, u_k$ at random.
- Get $\Sigma_{u_i, v}$ and Σ_{u_i, u_j} for all i, j .
- Compute and minimize

$$\sum_{i, j \leq k} \mathbb{1}_{\{u_i, u_j \text{ are not separated by } v\}}$$

centrality

The optimizing vertex satisfies

$$\mathbb{P} \left(c(\hat{v}) > \sqrt{\frac{11}{12}} n \right) \leq 2n \exp(-k/32).$$

This step only requires $\mathcal{O}(kn)$ queries.

If the maximum degree of T is d , then after removing $\hat{v}$, there are at most d connected components, all of them small.

One may identify the components using $\mathcal{O}(dn)$ queries.

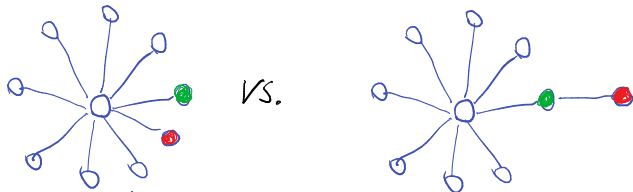
With probability $\geq 1 - \epsilon$, the algorithm recovers the tree
using

$$\mathcal{O} \left(n \log(n) \max \left\{ \log \left(\frac{n}{\epsilon} \right), d \right\} \right)$$

queries.

lower bound

$\Omega(dn)$ queries are necessary to recover trees with maximum degree d .



Impossible to tell them apart unless $\sum_{i=1}^n \mathbf{1}_{\{v_i \text{ is green}\}}$ is known. $\Omega(n^2)$ queries are necessary.

beyond trees—graphs of small treewidth

A much richer class of graphs is graphs with bounded **treewidth**.

Such graphs have small **balanced separators**. The key property is that

if $\text{treewidth} \leq k$ then G has a separator $S \subset V$ with $|S| \leq k + 1$ and

$$c(S) \leq \frac{1}{2} \frac{|V| - k}{|V| - (k + 1)}$$

where

$$c(S) = \frac{1}{|V \setminus S|} \max_{C \in \mathcal{C}^S} |C| ,$$

If we can find a small balanced separator S and identify the components of $V \setminus S$, then we could recurse similarly as before.

separators and rank

We say that S separates $A, B \subseteq V$ if any path from A to B contains a vertex in S .

Sullivant, Talaska, and Draisma (2010) proved that $\text{rank}(\Sigma_{A,B}) \leq r$ if and only if there is a set S with $|S| \leq r$ such that S separates A and B .

Thus,

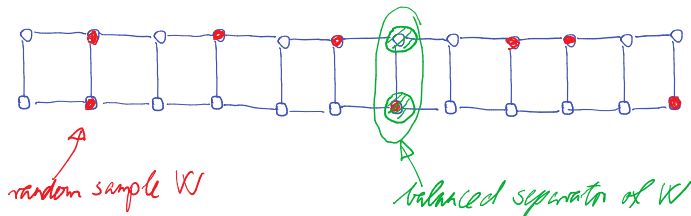
$$\text{rank}(\Sigma_{A,B}) \leq \min\{|S| : S \text{ separates } A \text{ and } B\}$$

and equality holds for generic Σ .

In particular, $\text{rank}(\Sigma_{AC,BC}) = \text{rank}(\Sigma_{A,B})$ if and only if C is a subset of a minimal separator of A and B .

finding a balanced separator

- Take a random sample W of m vertices of G . ($m = \mathcal{O}(k)$ suffices.)
- Find a set $S \subset V$ with $|S| \leq k + 1$ that is a balanced separator of W .
- Argue that S is a balanced separator of G as well.



finding a balanced separator of the random sample

Exhaustively search through the $2^{\mathcal{O}(k)}$ approximately balanced partitions $\mathbf{A}, \mathbf{B}$ of $\mathbf{W}$, that is, $\max(|\mathbf{A}|, |\mathbf{B}|) \leq (2/3)|\mathbf{W}|$.

For each $\mathbf{A}, \mathbf{B}$, find a minimal separator:

- determine all vertices $\mathbf{v} \in \mathbf{G}$ such that $\text{rank}(\Sigma_{\mathbf{A}, \mathbf{B}}) = \text{rank}(\Sigma_{\mathbf{A}\mathbf{v}, \mathbf{B}\mathbf{v}})$. (These are in some minimal separator.)
- Starting with $\mathbf{S} = \emptyset$, greedily add vertices to a set $\mathbf{S}$ for which $\text{rank}(\Sigma_{\mathbf{A}\mathbf{S}\mathbf{u}, \mathbf{B}\mathbf{S}\mathbf{u}}) = \Sigma_{\mathbf{A}, \mathbf{B}}$.

Next we argue that, with high probability, there is a good partition (almost balanced, minimal rank) and $\mathbf{S}$ is a balanced separator of $\mathbf{G}$ as well.

vc dimension

The argument hinges on Vapnik-Chervonenkis theory.

The small sample W is a good representative of the entire graph.

Lemma. (Feige and Mahdian, 2006.) For any graph G , let $\mathcal{F}_S$ be the set of all connected components of $G \setminus S$ and their complements in $V \setminus S$ and define

$$\mathcal{F}_k := \bigcup_{S: |S| \leq k} \mathcal{F}_S$$

Then the VC-dimension of $\mathcal{F}_k$ is at most $11k$.

The VC-inequality implies that for all sets $C \in \mathcal{F}_k$,

$$\frac{|C|}{|V|} - \delta \leq \frac{|W \cap C|}{|W|} \leq \frac{|C|}{|V|} + \delta.$$

whenever $|W| = \tilde{\Omega}(k/\delta^2)$.

connected components

Once a small balanced separator S is found, we determine the connected components of $G \setminus S$.

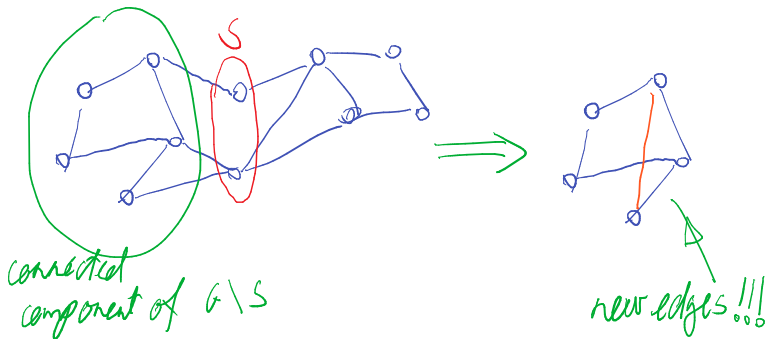
This can be done in linear time by checking ranks.

Bounded fragmentation is required. Similar to the case of trees.
Bounded degree suffices.

recursing in the connected components

Once the connected components of $G \setminus S$ are determined, we may recurse.

However, there is an extra complication.



conditional covariances

A basic property of Gaussian graphical models (Lauritzen, 1996):

If S separates C from the rest of G , then $K_C = (\Sigma_{C|S})^{-1}$.

We need to work with conditional covariances. These may be computed at a low cost using the ordinary covariance oracle.

main result

Let Σ be generic and let $G = \mathcal{G}(\Sigma)$ be such that $\text{tw}(G) \leq k$ and $\Delta(G) \leq d$. Then, with probability at least $1 - \epsilon$, the recursive algorithm reconstructs G with

$$\mathcal{O} \left((2^{\mathcal{O}(k \log k)} + dk \log n) k^2 n \log^3 n \log \frac{1}{\epsilon} \right),$$

queries.

The computational complexity is also similar.

We actually recover $K = \Sigma^{-1}$, not just the concentration graph.

further questions

- imperfect oracle.
- robustness.
- non-Gaussian models.
- property testing.