Learning Gaussian DAG Models without Condition Number Bounds

Constantinos Daskalakis, Vardis Kandiros, **Rui Yao**

June 16, 2025

What is a Directed Acyclic Graphical (DAG) Models?

Directed Acyclic Graph

Consider a directed acyalic graph (DAG) G, each node i associate with random variable X_i.

- Random variables in the DAG can be factorized along the DAG:

$$p(x) = \prod_{i=1}^{n} p(X_i | X_{\text{parents of } i}).$$

Structural Equation Model

In general, learning a DAG model is NP-hard!
A Structural Equation Model (SEM) guarantee the identifiability of the learning problem. [PMJS14, SHH+06]

Structural Equation Model

In general, learning a DAG model is NP-hard!
A Structural Equation Model (SEM) guarantee the identifiability of the learning problem. [PMJS14, SHH+06]

Linear SEM for Gaussian

Assume the conditional probability is linearly interpolated:

$$X_i = \sum_{j:j \text{ is a parent of } i} b_{ji} X_j + \varepsilon_i,$$

where b_{ji} are constant, ε_i are Gaussian Noise.

Background

Assumptions

1. Sparse graph: number of parents $\leq d$.
2. Noise ε_i are i.i.d. $\mathcal{N}(0,1)$.

The identifiability for this type of problems are established by [PB14], if the variances ε_i are equal.

Background

Assumptions

1. Sparse graph: number of parents $\leq d$.
2. Noise ε_i are i.i.d. $\mathcal{N}(0,1)$.

The identifiability for this type of problems are established by [PB14], if the variances ε_i are equal.

Prior work

1. General algorithm, such as PC algorithm, needs faithfulness assumption, which is strong.
2. Recent works [CDW19] and [GTA22], establishes $O(d \log n)$ samples, n is the number of vertices.
3. However, the sample complexity is a polynomial of κ, which is the condition number of the covariance matrix.

Our Goal

- We want to avoid condition number.
- Also we want to have an efficient algorithm

Our Goal

- We want to avoid condition number.
- Also we want to have an efficient algorithm

Value $\tau(G)$

$$\tau(G) = 1 + \max_i \sum_{j : i \text{ is a parent of } j} b_{ij}^2.$$

That is, 1 plus maximum of sum of square of the out-weight for every vertex.

Our Work

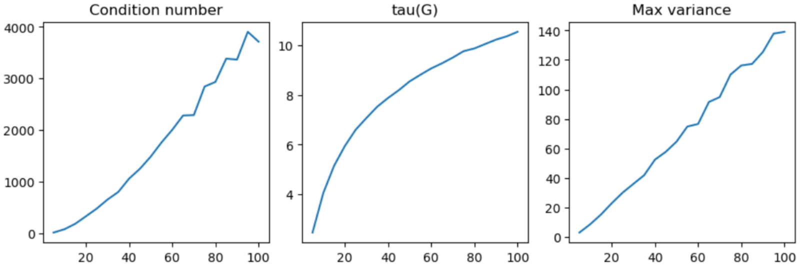

Figure 1: Comparison of the condition number, $\tau(G)$ and the max variance for a random graph.

Results

Inefficient algorithm. We can recover the graph with $O(\tau(G) \cdot d \log n)$ samples, in runtime $n^{O(d)}$.

Inefficient algorithm. We can recover the graph with $O(\tau(G) \cdot d \log n)$ samples, in runtime $n^{O(d)}$.

Lower bound. We cannot recover the graph with at most $O(\tau(G) \cdot \log n)$ samples.

Results

Inefficient algorithm. We can recover the graph with $O(\tau(G) \cdot d \log n)$ samples, in runtime $n^{O(d)}$.

Lower bound. We cannot recover the graph with at most $O(\tau(G) \cdot \log n)$ samples.

In addition to the algorithm with better sample complexity, we present our result for an efficient algorithm
Efficient algorithm. We can recover the graph with $O(\mathrm{poly}(\tau(G), d, R) \log n$ samples efficiently. Here, R is the maximum variance of all X_i.

Future Directions

- There is still a gap d between upper and lower bound. It is still a open problem even for undirected graphical model.
- Can we design an efficient algorithm that does not depend on R?
- How to generalize to other types of noises or unequal variance?

Thank you and have a nice day!

[CDW19] Wenyu Chen, Mathias Drton, and Y Samuel Wang.
On causal discovery with an equal-variance assumption.
Biometrika, 106(4):973–980, 2019.

[GTA22] Ming Gao, Wai Ming Tai, and Bryon Aragam.
Optimal estimation of gaussian dag models.
In *International Conference on Artificial Intelligence and Statistics*, pages 8738–8757. PMLR, 2022.

[PB14] Jonas Peters and Peter Bühlmann.
Identifiability of gaussian structural equation models with equal error variances.
Biometrika, 101(1):219–228, 2014.

[PMJS14] Jonas Peters, Joris M Mooij, Dominik Janzing, and Bernhard Schölkopf.
Causal discovery with continuous additive noise models.
2014.

[SHH+06] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan.
A linear non-gaussian acyclic model for causal discovery.
Journal of Machine Learning Research, 7(10), 2006.