Robust Identifiability in Linear Structural Equation Models of Causal Inference

Karthik Abinav Sankararaman and Anand Louis and Navin Goyal

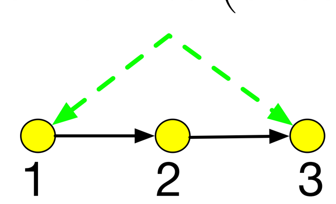
Meta AI, Austin and Indian Institute of Science, Bengaluru and Microsoft Research, India

Overview

- Study the **numerical stability** of LSEM parameter recovery via **condition number** on **general path graphs**
- A sufficient condition when parameter recovery problem is stable
- Random models satisfy the condition with substantial probability
- Experimental results

Linear Structural Equations (LSEM)

A **mixed** graph on the n (observable) variables.



- $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$ - matrix of edge weights of the DAG (strength of causal effect).
- $\mathbf{X} \in \mathbb{R}^{n \times 1}$ - random variables corresponding to the observable variables in the system with covariance $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$.
- $\eta \in \mathbb{R}^{n \times 1}$ - zero-mean Gaussian noises whose covariance matrix is $\mathbf{\Omega} \in \mathbb{R}^{n \times n}$.

LSEM assumes the following relationship between the random variables in $\mathbf{X}$.

$$\mathbf{X} = \mathbf{\Lambda}^T \mathbf{X} + \eta.$$

Gaussian assumption on η implies $\mathbf{X}$ is a multi-variate Gaussian with covariance

$$\mathbf{\Sigma} = (\mathbf{I} - \mathbf{\Lambda})^{-T} \, \mathbf{\Omega} \, (\mathbf{I} - \mathbf{\Lambda})^{-1}.$$

Typical setting. Experimenter estimates co-variance matrix $\mathbf{\Sigma}$ from finite samples, has a causal hypothesis represented as a mixed graph. Uses a parameter recovery algorithm, such as [2], to obtain the matrices $\mathbf{\Lambda}$ and $\mathbf{\Omega}$.

Challenges. Finite samples, noisy data to estimate $\mathbf{\Sigma}$. Recovery can potentially be bad. **We answer when can it be good?**

Condition number

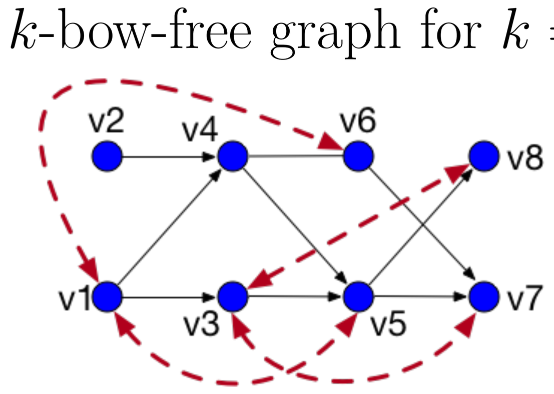
$$\kappa(\mathbf{\Lambda}, \mathbf{\Sigma}) := \lim_{\gamma \to 0^+} \sup_{\tilde{\mathbf{\Sigma}}_\gamma \in \mathcal{F}_\gamma} \frac{\mathrm{Rel}(\mathbf{\Lambda}, \tilde{\mathbf{\Lambda}}_\gamma)}{\mathrm{Rel}(\mathbf{\Sigma}, \tilde{\mathbf{\Sigma}}_\gamma)},$$

where

- $\mathrm{Rel}(\mathbf{A}, \mathbf{B}) := \max_{1 \le i \le n, 1 \le j \le m : |A_{i,j}| \neq 0} \frac{|A_{i,j} - B_{i,j}|}{|A_{i,j}|}$.
- $\mathcal{F}_\gamma$ - set of matrices $\tilde{\mathbf{\Sigma}}_\gamma$ such that $\mathrm{Rel}(\mathbf{\Sigma}, \tilde{\mathbf{\Sigma}}_\gamma) \le \gamma$.
- For any $\tilde{\mathbf{\Sigma}}_\gamma \in \mathcal{F}_\gamma$, $\tilde{\mathbf{\Lambda}}_\gamma$ is the corresponding recovered parameter

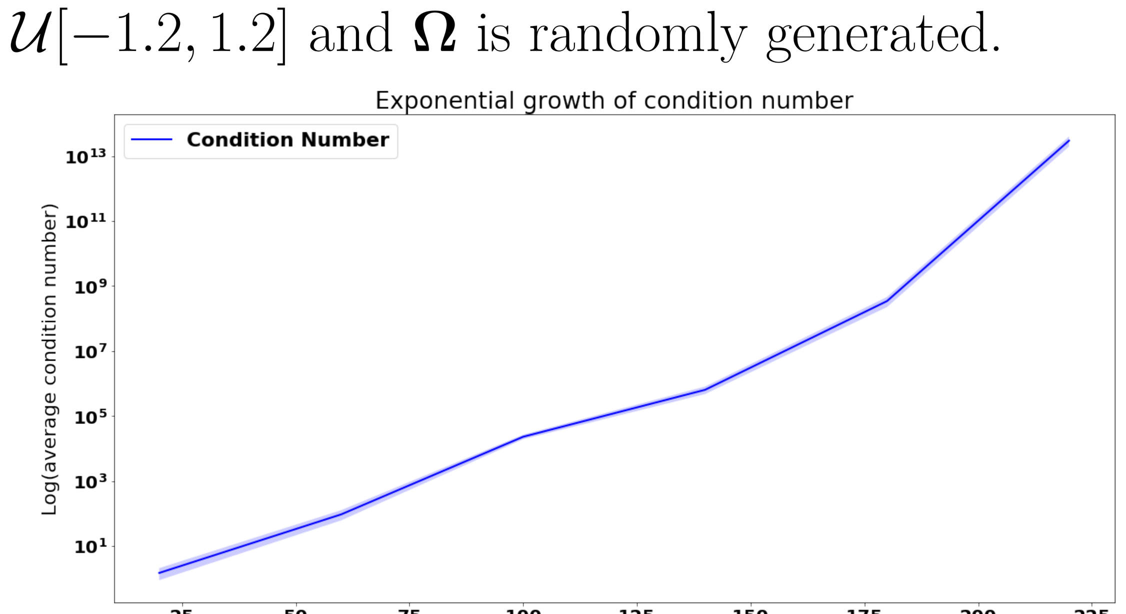
k-bow-free layered graph

The underlying DAG forms a directed layered graph and the mixed graph is bow-free [1]. Here we depict a k-bow-free graph for $k = 2$.



Exponential condition number

For every directed edge $i \to j$ we have $\lambda_{i,j} \sim \mathcal{U}[-1.2, 1.2]$ and $\mathbf{\Omega}$ is randomly generated.



Perturbation model

Let $\epsilon \in \mathbb{R}^{n \times n}$ denote the matrix of perturbations such that every entry (i, j) we have $\varepsilon_{i,j} \le \frac{\gamma}{\sqrt{k}} \Sigma_{i,j}$. Fix a small $0 < \gamma < \frac{1}{n^4}$. The perturbed matrix is $\tilde{\mathbf{\Sigma}} := \mathbf{\Sigma} + \epsilon$.

Sufficient condition for stability

$\mathbf{\Sigma} \succeq 0$ is an $n \times n$ symmetric matrix satisfying the following conditions for some $0 \le \alpha \le 1$. $\|\mathbf{\Sigma}_{[\mathrm{pa}(v),\, v]}\| \le \alpha \|\mathbf{\Sigma}_{[\mathrm{pa}(v),\, \mathrm{pa}(v)]}\|$, $\|\mathbf{\Sigma}_{[\mathrm{spa}(v),\, \mathrm{pa}(v)]}\| \le \alpha \|\mathbf{\Sigma}_{[\mathrm{pa}(v),\, \mathrm{pa}(v)]}\|$ and $\|\mathbf{\Sigma}_{[\mathrm{spa}(v),\, v]}\| \le \alpha \|\mathbf{\Sigma}_{[\mathrm{pa}(v),\, \mathrm{pa}(v)]}\|$ and $\kappa(\mathbf{\Sigma}_{[\mathrm{pa}(v),\, \mathrm{pa}(v)]}) \le \kappa_0 \le \frac{1}{2\gamma}$. The *true* parameter $\mathbf{\Lambda}$, corresponding to the $\mathbf{\Sigma}$, satisfies the following. For every i, j,

$$|\Lambda_{i,j}| \neq 0 \to \frac{1}{n^2} \le \frac{1}{\lambda} \le |\Lambda_{i,j}| \le \beta < 1.$$

If the above conditions hold, we have the following bound on the condition number for bow-free paths with n vertices.

$$\kappa(\mathbf{\Lambda}, \mathbf{\Sigma}) \le O(n^2).$$

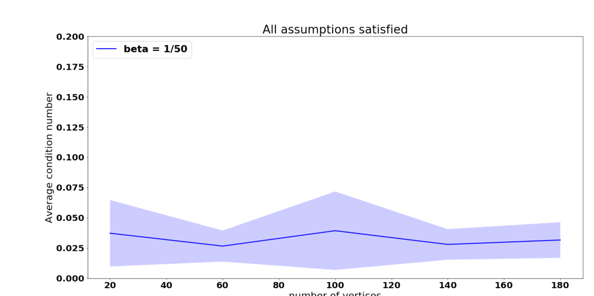
Condition satisified on $1 - \frac{1}{\mathtt{poly}(n)}$ measure of parameters

Generative Model.

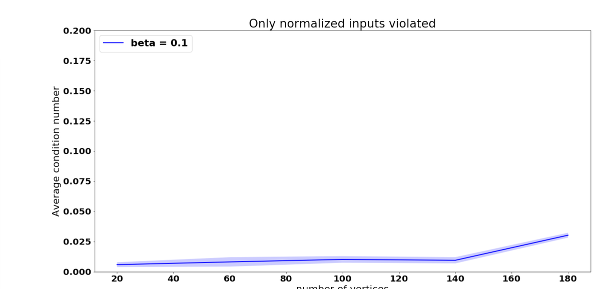
- $\mathbf{\Lambda} \sim \mathcal{U}\left[-\frac{1}{2k\mu}, \frac{1}{2k\mu}\right] \setminus \left[-\frac{1}{n^2}, \frac{1}{n^2}\right]$ when non-zero.
- Sample n-dimensional vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^d$ independently such that each vector $\mathbf{v}_i$ is a uniform sample from the sub-space perpendicular to $\mathtt{SPAN}(\{\mathbf{v}_j\}_{j \in V_{I-1}})$ and $\Omega_{i,j} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$. $V_I :=$ vertices in layer I.

When $\mu \ge 10(k+1)$ the generative model satisfies the sufficient condition with high-probability.

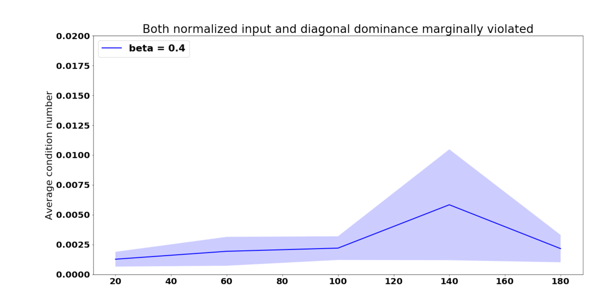
Experiments



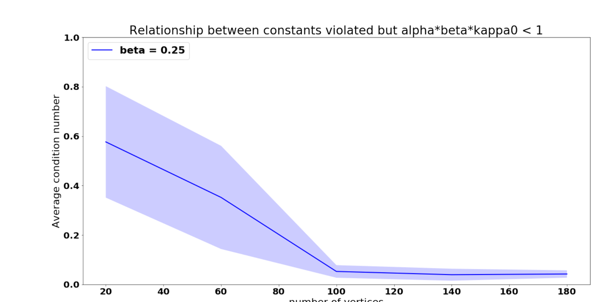
All assumptions satisfied



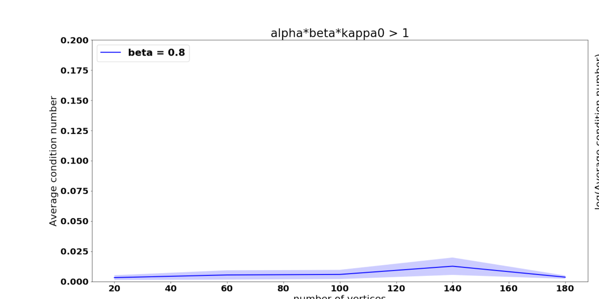
Only normalized input violated.



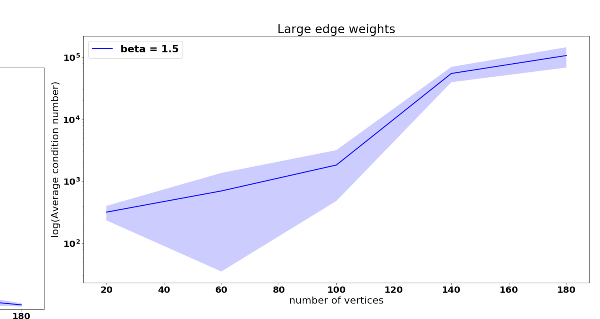
Normalized input and diagonal dominance marginally violated.



Relationship between the constants violated.



$\alpha * \beta * \kappa_0 > 1$.



$\beta > 1$ (y-axis is in log-scale).

Thus, only the requirement that $|\lambda_{i,j}| < 1$ is truly required assumption in practice. Rest are required to prove pessimistic theoretical bounds.

References

[1] Brito, C., and Pearl, J. A new identification condition for recursive models with correlated errors. *Structural Equation Modeling 9*, 4 (2002), 459–474.

[2] Foygel, R., Draisma, J., and Drton, M. Half-trek criterion for generic identifiability of linear structural equation models. *Annals of Statistics 40*, 3 (2012), 1682–1713.