Numerical Stability of Linear Structural Equation Models of Causality

Karthik Abinav Sankararaman and Anand Louis and Navin Goyal

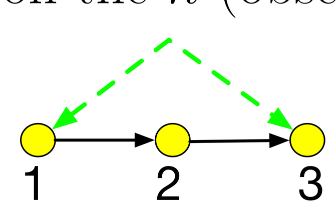
The University of Maryland, College Park and Indian Institute of Science, Bengaluru and Microsoft Research, India

Overview

- Study the **numerical stability** of LSEM parameter recovery via **condition number**
- A sufficient condition when parameter recovery problem is stable
- Random models satisfy the condition with substantial probability
- Exist examples that are extremely unstable
- 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 mumber

For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times m}$, relative distance is

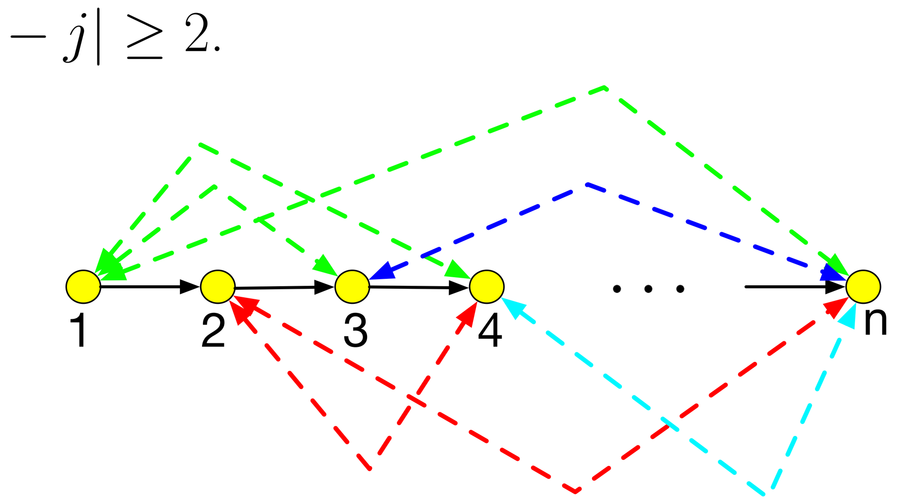
$$\mathrm{Rel}(\mathbf{A}, \mathbf{B}) := \max_{1\le i\le n, 1\le j\le m: |A_{i,j}|\neq 0} \frac{|A_{i,j} - B_{i,j}|}{|A_{i,j}|}.$$

ℓ_∞-condition number. Given $(\mathbf{\Sigma}, \mathbf{\Lambda})$. $\mathcal{F}_\gamma$ - set of matrices $\tilde{\mathbf{\Sigma}}_\gamma$ such that $\mathrm{Rel}(\mathbf{\Sigma}, \tilde{\mathbf{\Sigma}}_\gamma) \le \gamma$. For any $\tilde{\mathbf{\Sigma}}_\gamma \in \mathcal{F}_\gamma$, let the corresponding recovered parameter matrix be denoted by $\tilde{\mathbf{\Lambda}}_\gamma$. Then the relative ℓ_∞-condition number is defined as,

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

Bow-free path

The underlying DAG forms a directed path and the mixed graph is bow-free [1]. The bi-directed edges can exist between pairs of vertices (i, j) only if $|i - j| \ge 2$.



Perturbation model

Fix a $\gamma > 0$. For each $i, j \in [n]$, $\epsilon_{i,j} = \epsilon_{j,i}$ are arbitrary numbers satisfying $|\epsilon_{i,j}| = |\epsilon_{j,i}| \le \gamma\Sigma_{i,i}$.

$$\tilde{\Sigma}_{i,j} := \Sigma_{i,j} + \epsilon_{i,j} \qquad \forall (i, j).$$

Sufficient condition for stability

$\mathbf{\Sigma} \succeq 0$ is an $n \times n$ symmetric matrix satisfying the following conditions for some $0 \le \alpha \le \frac{1}{5}$.

$$|\Sigma_{i-1,i}|, |\Sigma_{i,i+1}|, |\Sigma_{i-1,i+1}| \le \alpha\Sigma_{i,i} \qquad \forall i \in [n-1]$$

The *true* parameter $\mathbf{\Lambda}$, corresponding to the $\mathbf{\Sigma}$, satisfies the following. For every i, j,

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

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

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

Random data satisfies the condition

Generative Model.

- $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$: each non-zero entry $\sim \mathcal{U}[-h, h]$.
- $\mathbf{\Omega} \in \mathbb{R}^{n\times n}$: Sample n-dimensional vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^d$ independently from a d-dimensional unit sphere, such that $\langle \mathbf{v}_i, \mathbf{v}_{i-1}\rangle = 0$. $\Omega_{i,j} = \langle \mathbf{v}_i, \mathbf{v}_j\rangle$.

$h < 1$ implies the above generative model satisfies the sufficient condition with high-probability.

Instability example

Graph on 4 vertices.

$$\mathbf{\Omega} = \begin{bmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 1/2 \\ 1/2 & 0 & 1 & 0 \\ 1/2 & 1/2 & 0 & 1 \end{bmatrix} \succeq 0$$

$\Lambda_{1,2} = \sqrt{2}, \Lambda_{2,3} = -\sqrt{2}, \Lambda_{3,4} = \frac{1}{2}$.
After perturbation $\tilde{\Lambda}_{3,4} = 1$.
Therefore, $\mathrm{Rel}(\mathbf{\Lambda}, \tilde{\mathbf{\Lambda}}) = O(1) \neq 0$.

Experiments

Figure 1: Layered graph: $p = 0.2$.

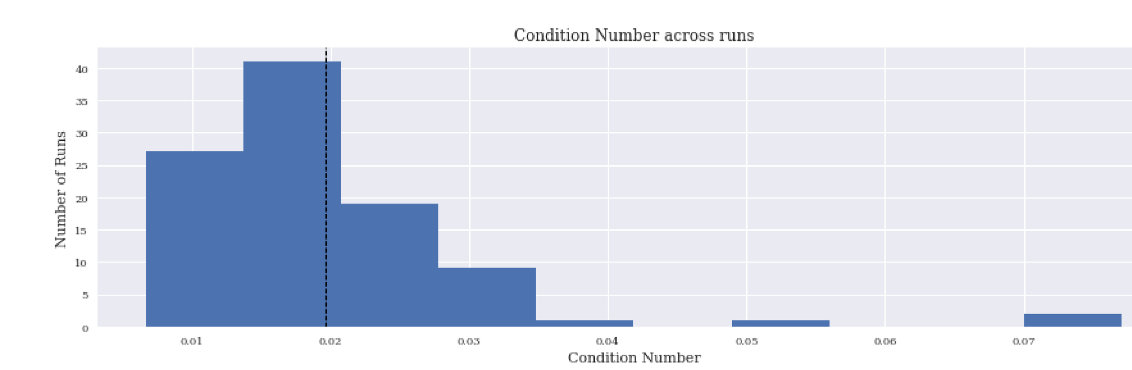


Figure 2: Clique of paths with $n = 30$ and $k = 5$.

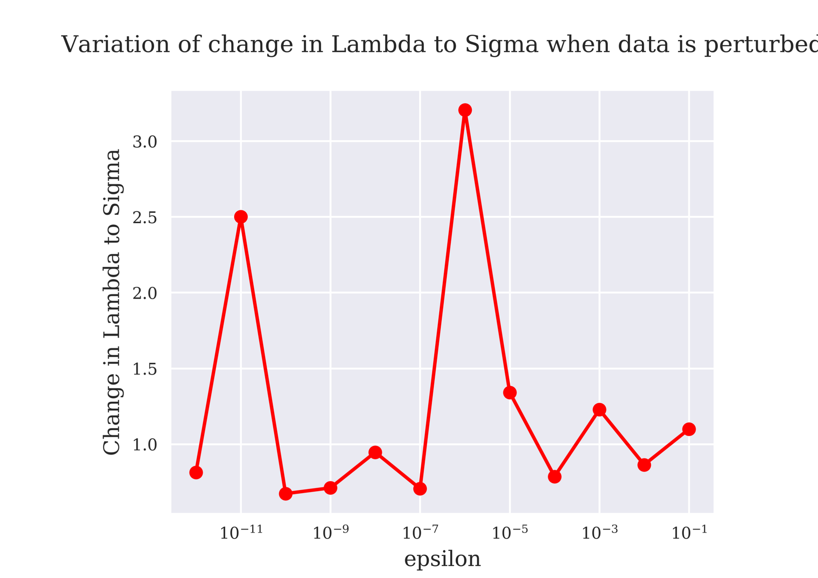


Figure 3: Variation of the randomized condition number as a function of the perturbation error on the sociology dataset

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.