Causal Discovery in Linear Latent Variable Models Subject to Measurement Error

Yuqin Yang[1] AmirEmad Ghassami[2] Mohamed Nafea[3] Negar Kiyavash[4] Kun Zhang[5,6] Ilya Shpitser[2]

[1]Georgia Tech [2]Johns Hopkins University [3]University of Detroit Mercy [4]EPFL [5]Carnegie Mellon University [6]MBZUAI

Motivation

There are complexities in real-life data that make causal discovery challenging

- Two main sources of complexities: **Latent confounding** and **measurement error**

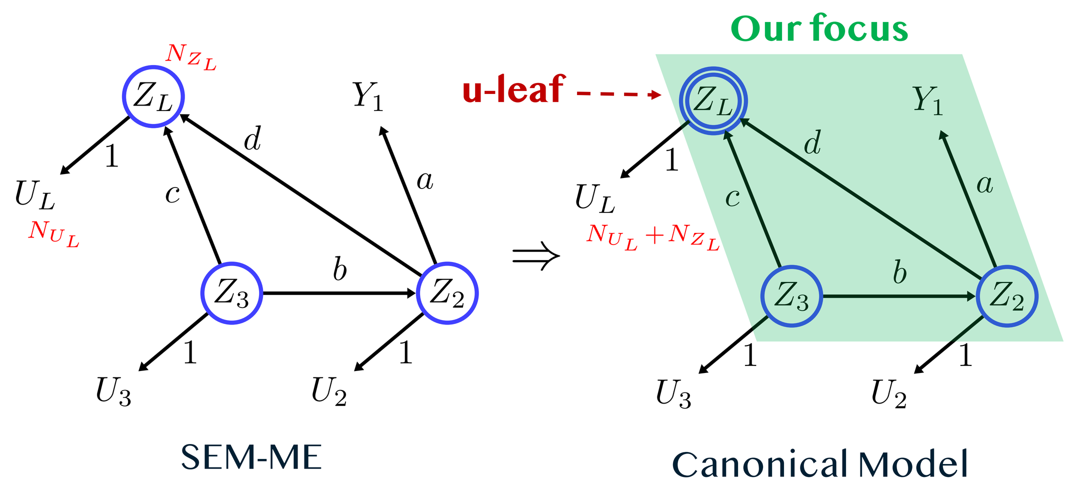
Majority of causal discovery methods assume that there are no such complexities in the system

- Leading to incorrect recovery on real data

Model Definition

Linear SEM with Measurement Error (SEM-ME)

- Underlying model: $V = \mathbf{C}V + N_V$
- V can be partitioned into $[Z; Y]$
  - Y: Observed variables
    Measured without error
  - Z: Unobserved variables
    Measured with error: $U_i = Z_i + N_{U_i}$
- **Canonical form**: Without loss of generality, unobserved leaf nodes (**u-leaf** nodes) are assumed to have no exog. noise terms [1]



Linear SEM with Unobserved Roots (SEM-UR)

$$H = N_H, \quad X = \mathbf{B}H + \mathbf{A}X + N_X$$

Latent variable · Observed variable · Adjacency matrix

- Assuming latent variables to be roots does not affect the estimated total causal effects among observed variables

Assumption 1 (Separability): Mixing matrix transforming exog. noises to obs. variables can be recovered from obs. distribution

- Satisfied when all noises are non-Gaussian

Mapping Between ME & UR Model

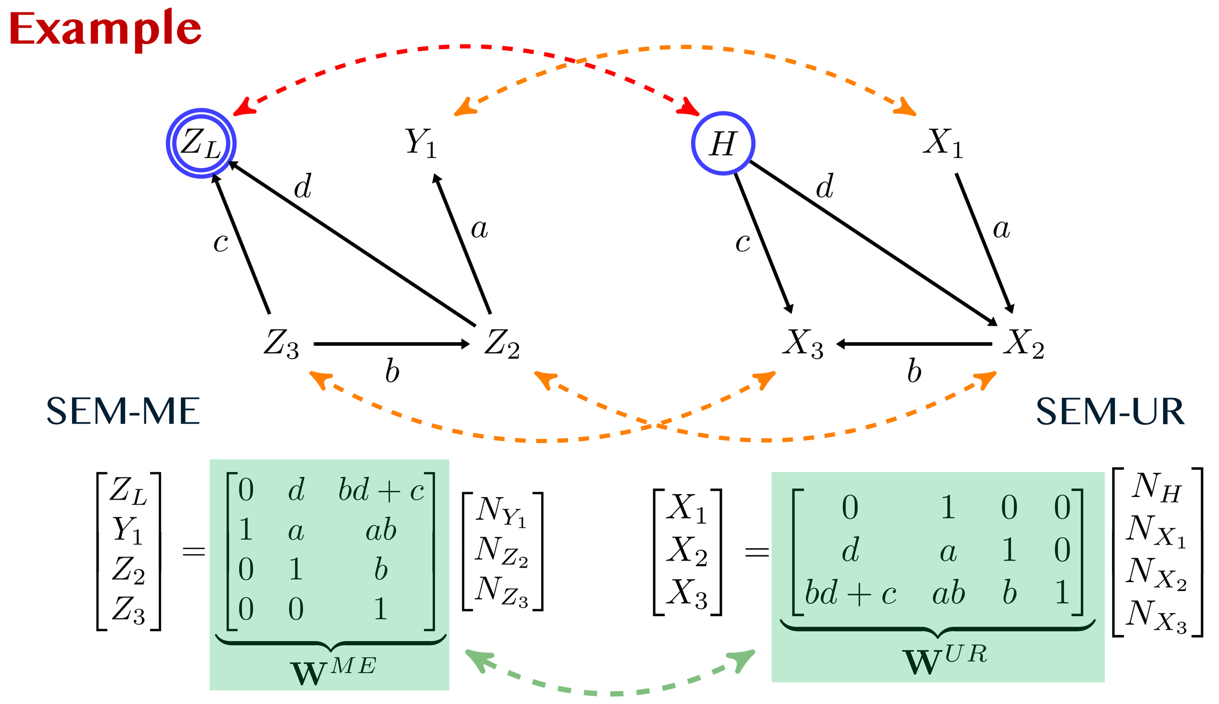
Mapping between the weighted causal diagrams of a SEM-ME and a SEM-UR:

- u-leaf ↔ latent
- non u-leaf ↔ observed
- Reverse all edges

Theorem 1: The mixing matrix of the models under the mapping are transpose of one another.

Remark: Any identifiability result based on mixing matrix for one model can be translated to the other model.

Example



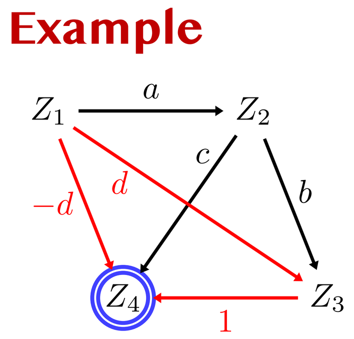
Identifiability Results

We characterize the extent of identifiability for both models under mild assumptions.

Assumption 2 (Two-fold faithfulness)

- Part (a): Conventional faithfulness in linear models
- Part (b): Prevents certain measure-zero parameter cancellation or proportionality
  - Strictly weaker than bottleneck faithfulness [2]

Example



- $Z_4 = (b + c)Z_2 + N_{Z_3}$
- The causal effect of Z_1 and Z_2 on Z_4 can be summarized by Z_2 alone, due to the parameter cancellation on the red triangle

Ancestral Ordered Grouping (AOG) and Direct Ordered Grouping (DOG)

Variables are partitioned into distinct groups such that:

- Each group contains at most one non u-leaf node
- Graph induced on each group is a star graph
- u-leaf nodes are assigned either to the group of a parent or a separate group based on different graphical conditions (see the paper)
- DOG is a finer partition than AOG

Identifiability Results (Cont'd)

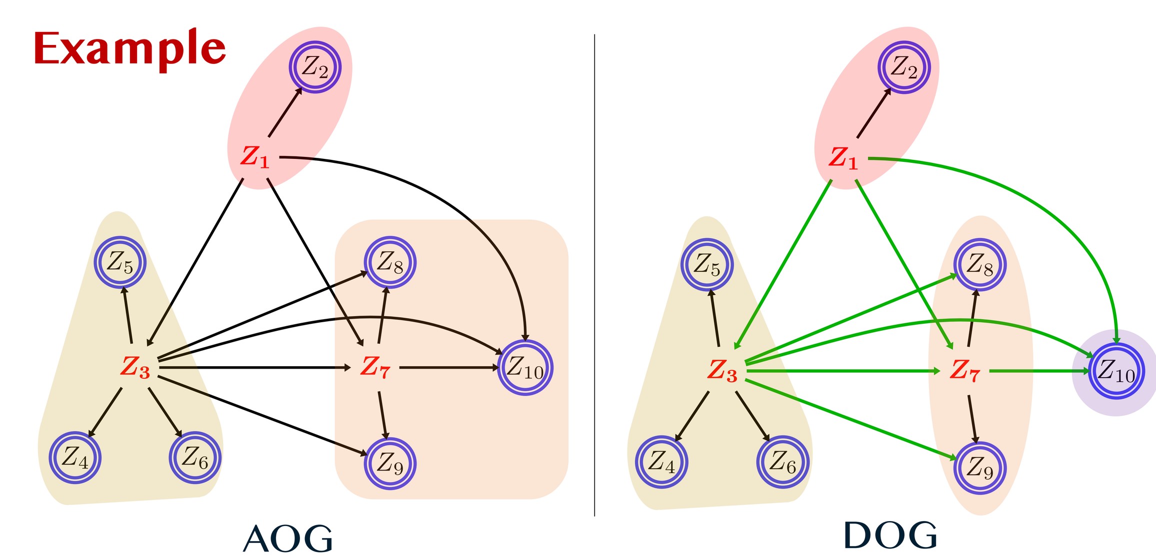
Theorem 2: Under Assumptions 1 & 2(a), a SEM-ME & a SEM-UR can be recovered up to its **AOG Equivalence Class (AOG-EC)**.

Theorem 3: Under Assumptions 1 & 2, a SEM-ME & a SEM-UR can be recovered up to its **DOG Equivalence Class (DOG-EC)**.

Recovery	AOG-EC	DOG-EC
(i) Order among groups	Yes	Yes
(ii) Edges across groups	No	Yes
(iii) Center of each group	No	No

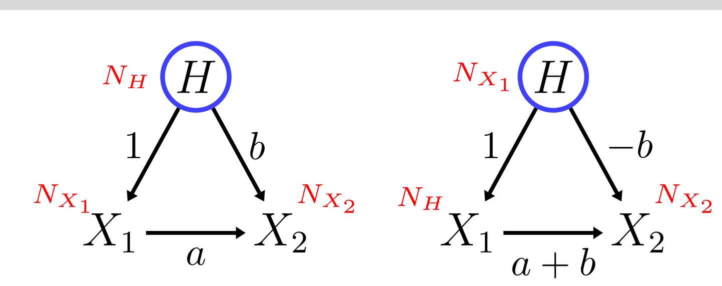
Remark: Model can be identified by the choice of the centers of the stars (or their corresponding exogenous noise terms) in each group.

Example



- Edges across groups can be identified for DOG, bot not for AOG.

Corollary: The **structure** of a SEM-UR can be **uniquely** identified.



DOG Recovery Algorithm

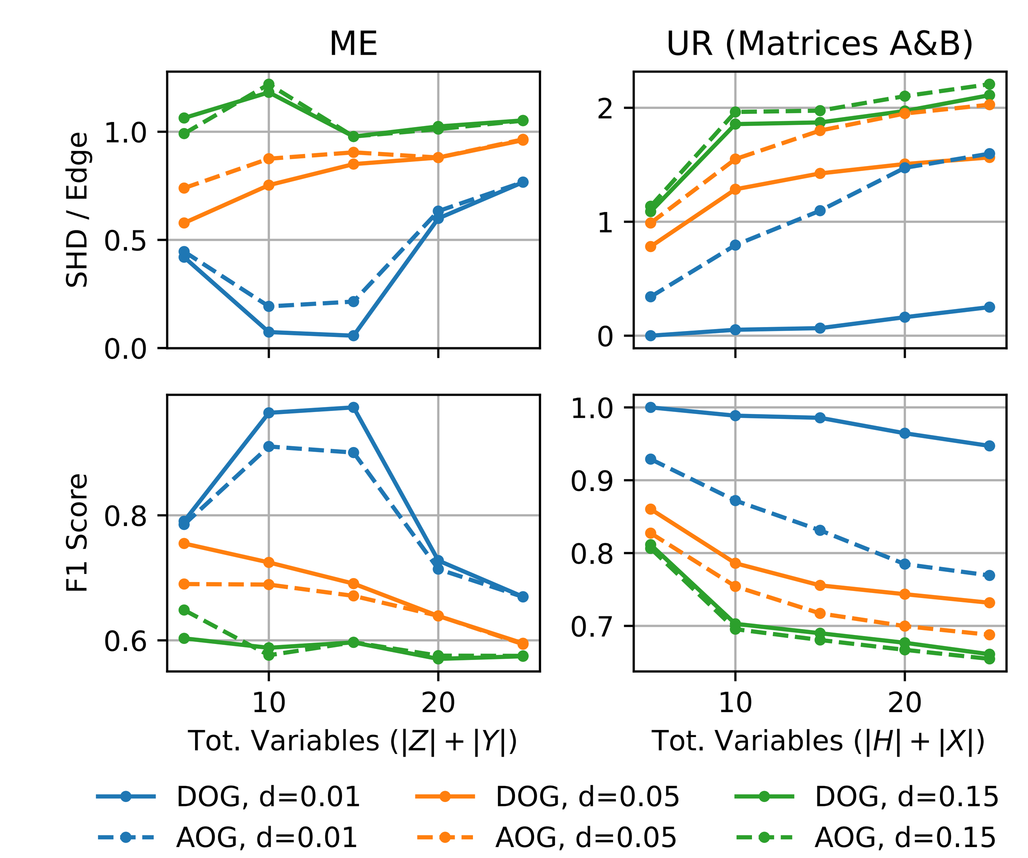
Proposition: Under Assumptions 1 & 2, any model that belongs to the AOG-EC of the ground-truth but not the DOG-EC has **strictly** more edges.

1) Recover the mixing matrix from observed data.
2) Return the AOG of the model by checking the support of the mixing matrix. (See the paper)
3) For all possible choices of the centers, find a choice that leads to the graph with fewest number of edges in the recovered model.

Simulations

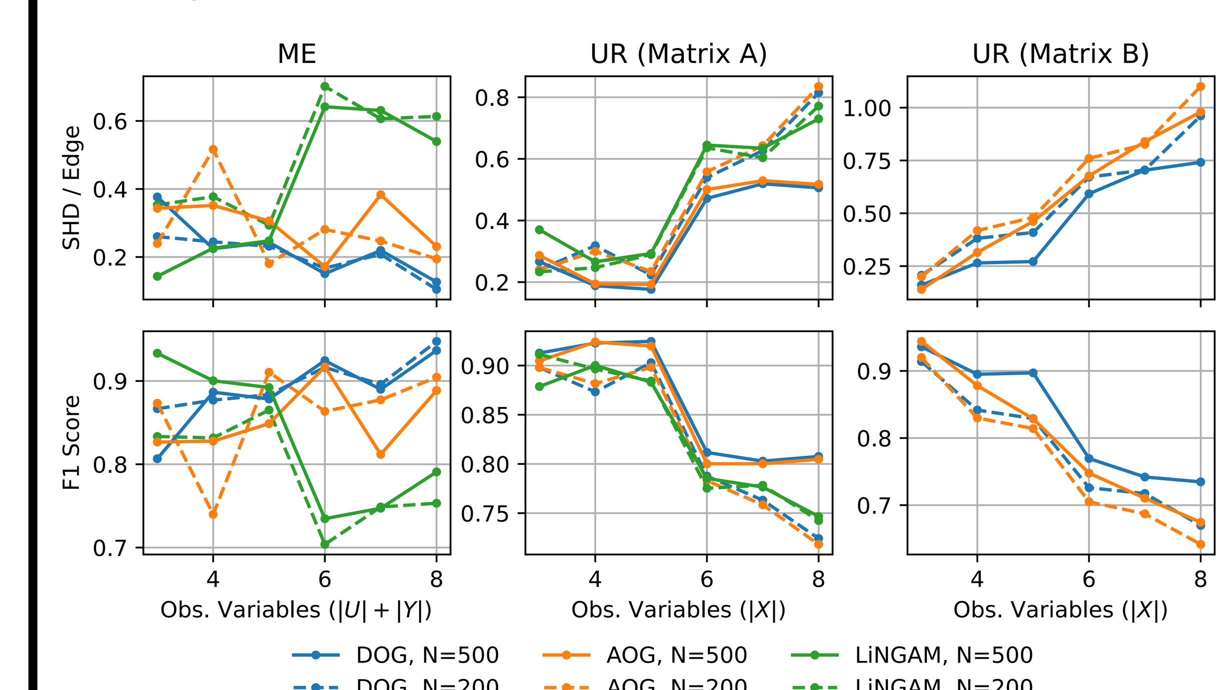
We compare the performance of our DOG recovery algorithm with AOG-based algorithm [3] and LiNGAM on both models under two settings.

(1) A noisy version of the true mixing matrix is given as input



- d: Variance of the added Gaussian noise on the mixing matrix

(2) Synthetic data with non-Gaussian noise



- N: Sample size / Number of obs. variables
- Estimate the mixing matrix from data using Reconstruction ICA

Results show that our algorithm outperforms AOG-based method, and both outperforms LiNGAM.

This demonstrates the necessity of using methods designed specifically to handle complexities.

References

[1] Zhang et al. "Causal Discovery with Linear Non-Gaussian Models under Measurement Error: Structural Identifiability Results." *UAI. 2018.*
[2] Adams et al. "Identification of Partially Observed Linear Causal Models: Graphical Conditions for the Non-Gaussian and Heterogeneous Cases." *NeurIPS. 2021.*
[3] Salehkaleybar et al. "Learning Linear Non-Gaussian Causal Models in the Presence of Latent Variables." *JMLR. 2020.*