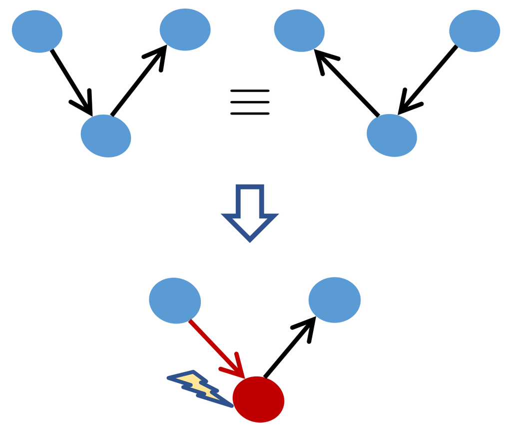
# Scalable Intervention Target Estimation in Linear Models

Burak Varıcı  Ali Tajer  Karthikeyan Shanmugam  Prasanna Sattigeri

Rensselaer Polytechnic Institute  IBM Research AI

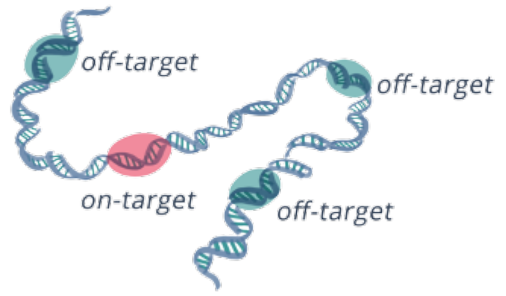

## Motivation

- Directed Acyclic Graphs (DAG): encode cause-effects
- Structure learning: up to Markov Equivalence Class (MEC)
- **Intervention**: Forcible changes to target variables



**Why estimate intervention targets?**

Interventional Structure Learning:
- Known targets → strong assumption
- Unknown targets → not scalable methods




Courtesy of SJTU-BioX

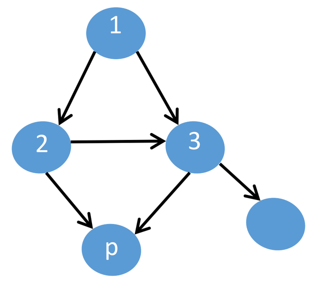
Stand-alone importance:
- Difference estimation → identify hub nodes
- Large-scale cloud systems → fault localization

Key challenges: computational complexity, restrictive assumptions, sample complexity

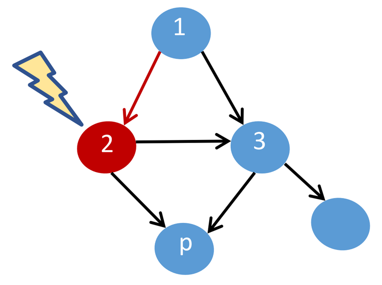
## Model

$\mathcal{G}_1 \triangleq (B^{(1)}, \epsilon^{(1)})$

$(\Sigma^{(1)}, \Theta^{(1)}, \Omega^{(1)})$




Observational




Interventional

$\mathcal{G}_2 \triangleq (B^{(2)}, \epsilon^{(2)})$

$(\Sigma^{(2)}, \Theta^{(2)}, \Omega^{(2)})$

- **Linear Structural Equation Model**: $X = [X_1, \dots, X_p]^\top$, and $\epsilon \sim (N, \Omega)$

$$X = B^\top X + \epsilon$$

- Precision matrix: $\Theta = (I - B)\Omega^{-1}(I - B)^\top$.
- **Soft Interventions:** Change in noise variations of targets.

$$\mathcal{I} \triangleq \{i : \sigma_i^{(1)} \neq \sigma_i^{(2)}\}$$

- Non-intervened parents: $\text{pa}(\mathcal{I}) \triangleq \{\text{pa(i)} \setminus \mathcal{I} : \text{i} \in \mathcal{I}\}$

## Problem Statement

Objective: estimate $\mathcal{I}$ and $\text{pa}(\mathcal{I})$ from $\Sigma^{(1)}$ and $\Sigma^{(2)}$

$$\max_{\hat{\mathcal{I}} \subseteq [p]} \mathbb{P}(\mathcal{I} = \hat{\mathcal{I}}) \quad \text{and} \quad \max_{\text{pa}(\hat{\mathcal{I}}) \in [p] \times [p]} \mathbb{P}(\text{pa}(\mathcal{I}) = \text{pa}(\hat{\mathcal{I}}))$$

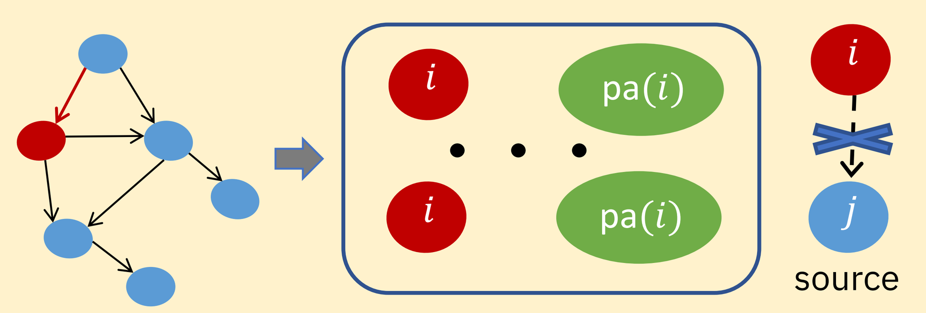
## Algorithmic Framework

- Marginal SEMs: $X_S \to (B_S, \epsilon_S)$
- Precision Difference Estimation (PDE): $\Delta_{\Theta_S} = \Theta_S^{(1)} - \Theta_S^{(2)}$.
- Lasso formulation and solution through ADMM (Jiang et al. JMLR 2018).

$$\hat{\Delta}_\Theta = \min_{\Delta_\Theta} \left\{ \frac{1}{2} \text{Tr}(\Delta_\Theta^\top \hat{\Sigma}^{(1)} \Delta_\Theta \hat{\Sigma}^{(2)}) - \text{Tr}(\Delta_\Theta(\hat{\Sigma}^{(1)} - \hat{\Sigma}^{(2)})) + \lambda \|\Delta_\Theta\|_1 \right\}$$

- Intervened $i$ : $\epsilon_S$ is never invariant
- Non-intervened $j$ : intervened ancestors and their parents $\text{pa}^+(\text{an}_{\mathcal{I}}(j))$ ✓
- **Goal** : efficiently search for an $S$ to make $[\Delta_{\Theta_S}]_{j,j} = 0$.

Pruning

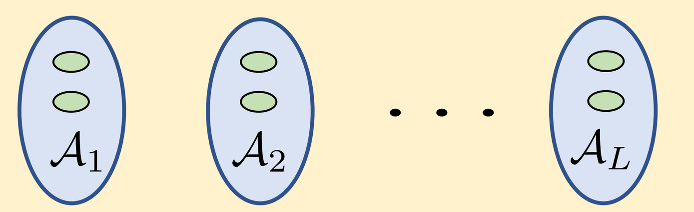



# PDE $= p + 1$

$S_\Delta \triangleq \{k : [\Delta_\Theta]_{k,k} \neq 0\} = \mathcal{I} \cup \text{pa}(\mathcal{I})$
(all nodes of interests)

$J_0 \triangleq \{j : j \in S_\Delta,\ j \notin \mathcal{I},\ \text{an}_{\mathcal{I}}(j) = \emptyset\}$
(non-intervened source nodes)

Grouping



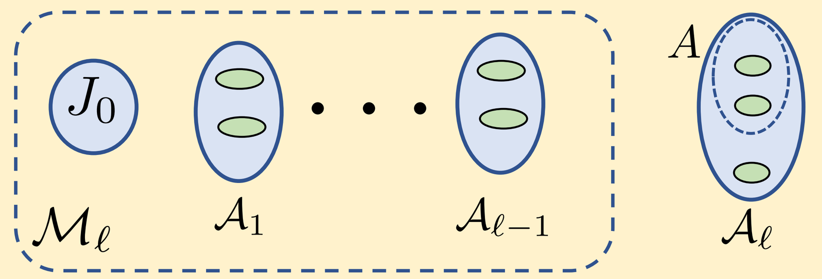

# PDE $= O(|S_\Delta|^2)$

Equivalence classes: $S_\Delta \setminus J_0 = \bigcup_{\ell \in [L]} \mathcal{A}_\ell$

$\mathcal{A}_\ell$ : same source ancestral set

Topological ordering: $\mathcal{A}_1 > \cdots > \mathcal{A}_L$

Identification




# PDE $= O(2^{|\mathcal{A}_\ell|}) << O(2^{|S_\Delta|})$

$\forall\, j \in \mathcal{A}_\ell \setminus \mathcal{I}$, guaranteed to find $A \subset \mathcal{A}_\ell$

$$[\Delta_{\Theta_{\mathcal{M}_\ell \cup A}}]_{j,j} = 0$$

Estimate $\Delta_{\Theta_{\mathcal{M}_\ell \cup A}}$ **only for each** $A \subseteq \mathcal{A}_\ell$.

## Consistency Guarantees

Theorem 1 & 2 (Consistency): Given the true covariances $\Sigma^{(1)}$ and $\Sigma^{(2)}$, the algorithm estimates:

- $\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I}) = 1$ : scalable since $2^{\max|\mathcal{A}_\ell|} << 2^{|S_\Delta|}$
- $\mathbb{P}(\text{pa}(\hat{\mathcal{I}}) = \text{pa}(\mathcal{I})) = 1$ : infers all the interventional information
- Refines MEC to $\mathcal{I}$-MEC : modularity with observational algorithms

## Sample Complexity

Theorem 3 (Sample complexity): Given the condition number of the matrix estimation problem is bounded, the algorithm has guarantees

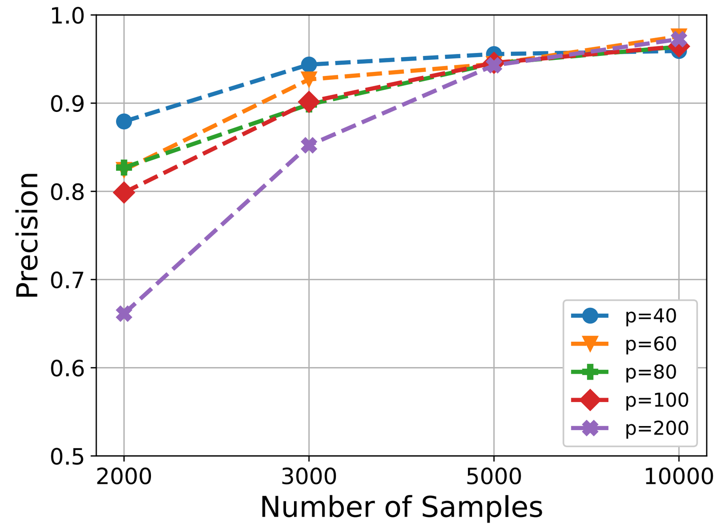
$$\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I}) \geq 1 - \delta \quad \text{and} \quad \mathbb{P}(\text{pa}(\hat{\mathcal{I}}) = \text{pa}(\mathcal{I})) \geq 1 - \delta$$

with number of samples $n = O(\frac{d^4 \log p}{\delta})$

✓ finite-sample guarantees
✓ sample efficiency

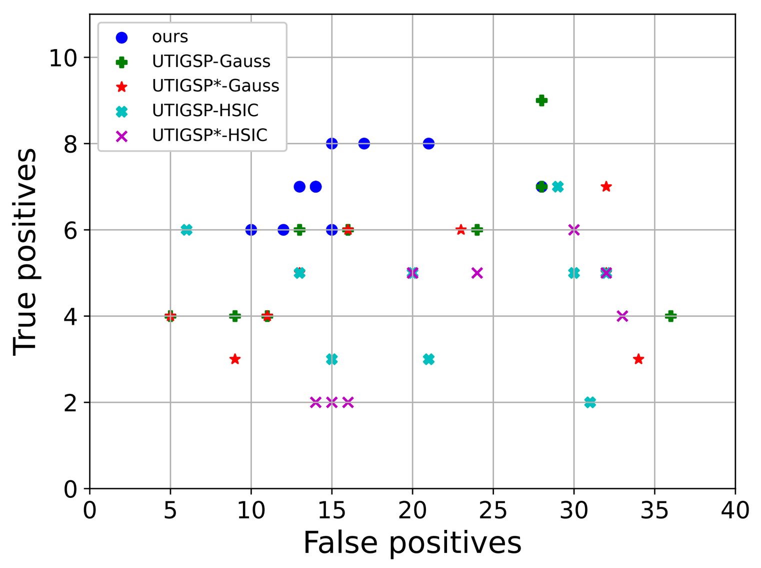
## Experiments

- **Synthetic data**: Generate Erdös-Renyi random DAGs with various graph sizes $p$.
- Compared to state-of-the-art algorithm: UT-IGSP (Squires et al., UAI 2020)
- **Scalability**: Runtime grows gracefully.




| | UT-IGSP | | Ours | |
|---|---|---|---|---|
| p | F1 | Time(s) | F1 | Time(s) |
| 40 | 0.99 | 0.8 | 0.95 | 0.1 |
| 60 | 0.97 | 5.2 | 0.96 | 0.2 |
| 80 | 0.97 | 17.8 | 0.96 | 0.3 |
| 100 | 0.96 | 61.2 | 0.96 | 0.3 |

- **Real data**: Protein signaling network (Sachs et al. 2005).




- DAG estimate: Combine recovered edges.
- ROC curves: Run PDEs with different $\lambda$'s.
- Outperforms UT-IGSP variants.

## Conclusions

- **Intervention target estimation** in linear models.
- Leveraged precision differences and proposed a **consistent** and **scalable** algorithm. Furthermore, refined the Markov equivalence class.
- Established finite-sample results for Gaussian models.