Coarsening causal DAG models

Alex Markham

Department of Mathematical Sciences, University of Copenhagen

EuroCIM'26

Joint work with...

Francisco Madaleno

Technical University of Denmark

Pratik Misra

Binghamton University

Overview

- Problem:
  - high-dimensional causal discovery is hard
  - observed features aren't necessarily at desired casual scale
- Our approach:
  - learn coarsened causal DAG, directly from data rather than from fine-grained DAG over observed features
  - describe search space as sublattice of partition refinement lattice, leading to efficient search
- Results:
  - lower sample and time complexity than learning fine-grained DAG
  - framework for understanding existing causal discovery algorithms and developing new ones

Theory

General coarsening

Definition
Given a DAG $G = (V, E)$, a *coarsening* is a DAG $G' = (V', E')$ for which there exists a surjection $\chi : V \to V'$ such that:

$$E' = \{\chi(v) \to \chi(w) \mid v \to w \in E, \chi(v) \neq \chi(w)\}.$$

Lemma
Let G be a DAG and G' be one of its coarsenings. Every distribution Markov to G is also Markov to G', that is, $\mathcal{M}(G) \subseteq \mathcal{M}(G')$.

Theorem
For any DAG $G = ([d], E)$, the poset of its coarsenings is a lattice, more specifically, a sublattice of the partition refinement lattice of $[d]$.

Example

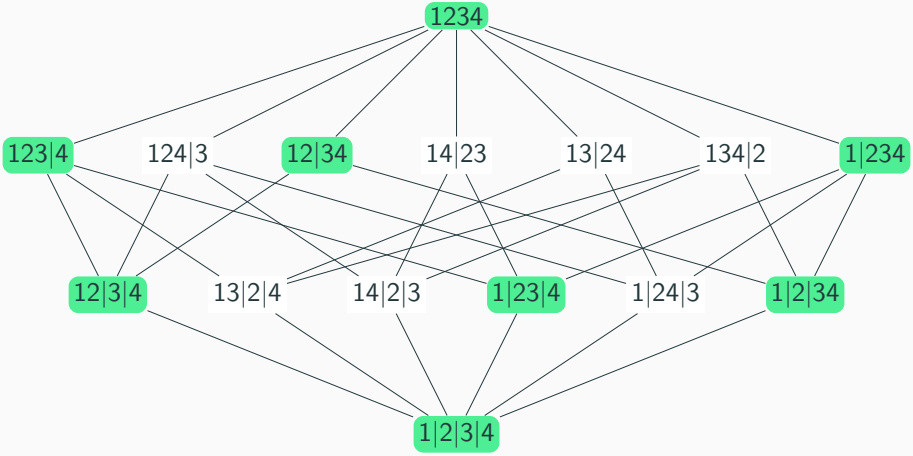


Figure 1: The partition refinement lattice on 4 nodes, with an example coarsening sublattice highlighted, for ground-truth $\underline{G} : 1 \to 2 \to 3 \to 4$.

A theoretical algorithm

High-level algorithm: `RePaRe` starts at trivial coarsening $\overline{G} = ([d], \emptyset)$ and recursively:

- uses `Refine` to pick a part and split it
- uses `IsEdge` to add directed edges between new and existing parts

Completeness: `RePaRe` can construct any valid coarsening by traversing the lattice

Identification (estimation) of any particular coarsening becomes a matter of appropriately defining it, `Refine`, and `IsEdge`

A framework for causal discovery algorithms generally

A practical algorithm

Interventional coarsening

Motivation: given a collection of post-intervention samples:

- which of the measured features are affected similarly?
- what is the salient causal structure among these clusters?

Approach: formalize "measured features being affected similarly":

$$\mathcal{I}\text{-an}_G(v) \coloneqq \{w \in \text{an}_G(v) \mid \exists I \in \mathcal{I} \text{ such that } w \in I\}.$$

Definition

Given DAG $G = (V, E)$ and set of intervention targets $\mathcal{I}$, the interventional coarsening, denoted $G^{\mathcal{I}}$, defined by partition surjection χ satisfying

$$\text{for all } v, w \in V, \quad \chi(v) = \chi(w) \iff \mathcal{I}\text{-an}_G(v) = \mathcal{I}\text{-an}_G(w)\,.$$

Interventional coarsening

Assumption

Given a ground-truth DAG G and intervention set $\mathcal{I}$:

1. *coarse interventional Markov: $\chi(v)$ independent of non-descendants given parents; non-intervened factors remain invariant*
2. *coarse faithfulness: the only conditional independences among coarse nodes $\chi(V)$ are those implied by the Markov property.*
3. *interventional soundness: interventions induce measurable changes in marginal distributions compared to the observational distribution*

Theorem

Interventional coarsening $G^{\mathcal{I}}$ is identifiable from unlabeled interventional distributions.

A practical algorithm

- assume linear Gaussian ground truth DAG model
- replace `Refine` with a t-test to compute $\mathcal{I}\text{-an}_G()$
- replace `IsEdge` with CCA and Wilks' Λ
- time complexity is $O\big(den + k^2p^2n\big)$
- can relax linear Gaussian assumptions by:
  - using energy distance or MMD for `Refine`
  - using distance correlation or HSIC for `IsEdge`
  - at the cost of increased time complexity

Experiments

Synthetic data

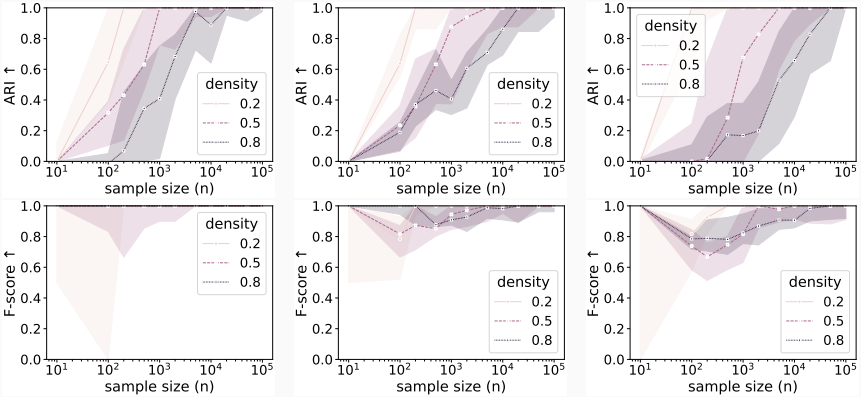


Figure 2: Convergence. **From left to right**: Varying intervention budgets ($\iota = 2, 5, 8$).

Synthetic data

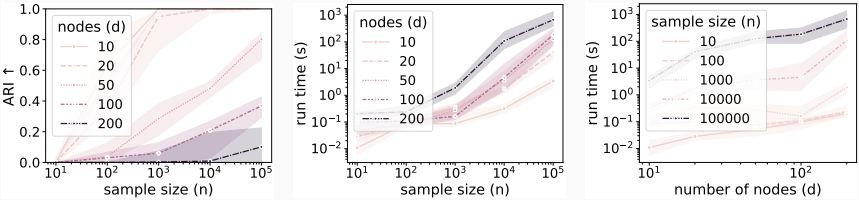
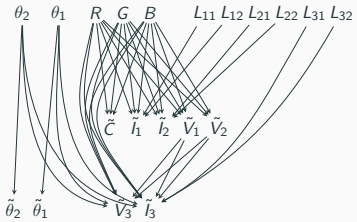
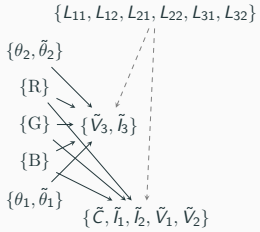


Figure 3: Scalability on synthetic data, with density 0.2 and $\iota = 5$.

Real data



(a) Ground-truth light-tunnel DAG.



(b) Coarsening learned by RePaRefrom interventional data.

Real data

Method (Selection)	ARI	Precision	Recall	F-score	Run time (s)
GIES (score)	–	0.632	0.615	0.623	7.227
GnIES (score)	–	0.426	0.513	0.465	1966.916
UT-IGSP (score)	–	0.333	0.385	0.357	1.117
RePaRe(grouped, score)	0.932	1.000	0.500	0.667	0.063
RePaRe(grouped, optimal)	1.000	1.000	0.500	0.667	0.141
RePaRe(ungrouped, score)	1.000	1.000	0.800	0.889	0.753
RePaRe(ungrouped, optimal)	1.000	1.000	0.800	0.889	0.753

Table 1: Light-tunnel real data results, RePaRe compared to baselines.

Conclusion

Summary

- Problem:
  - high-dimensional causal discovery is hard
  - observed features aren't necessarily at desired casual scale
- Our approach:
  - learn coarsened causal DAG, directly from data rather than from fine-grained DAG over observed features
  - describe search space as sublattice of partition refinement lattice
- Results:
  - lower sample and time complexity than learning fine-grained DAG
  - framework for understanding existing causal discovery algorithms and developing new ones

Thanks for listening!

- Published at Conference on Causal Learning and Reasoning (CLeaR)
- Read the preprint: arXiv:2601.10531
- Email me to discuss: awm@math.ku.dk
- See the code: github.com/Alex-Markham/repare

Questions?