



UNIVERSITÄT ZU LÜBECK



KØBENHAVNS
UNIVERSITET

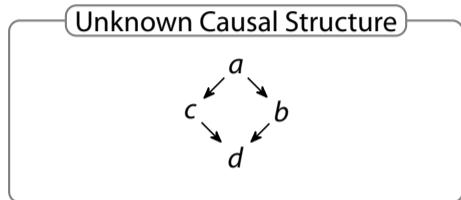
Embracing Discrete Search: A Reasonable Approach to Causal Structure Learning

Marcel Wienöbst, Leonard Henckel, Sebastian Weichwald

ICLR 2026

Causal Discovery

Sampling Process



Observations

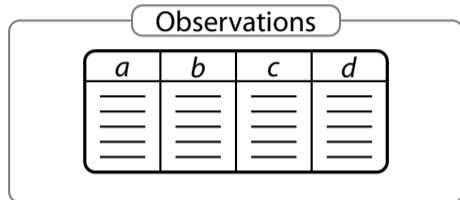
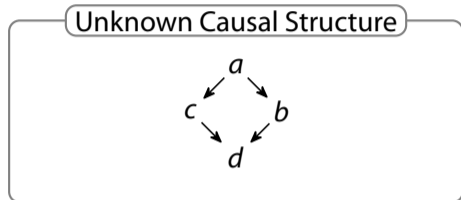
a	b	c	d
—	—	—	—
—	—	—	—
—	—	—	—
—	—	—	—



?

Causal Discovery

Sampling Process



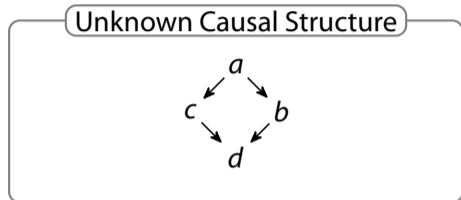
?

Assumptions

- Causal sufficiency
- Acyclicity
- Faithfulness
- Linear additive noise models

Causal Discovery

Sampling Process



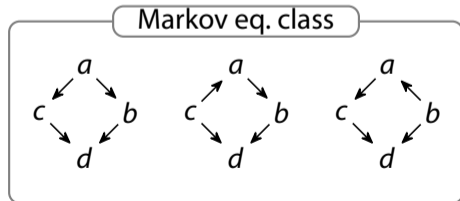
Observations

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
—	—	—	—
—	—	—	—
—	—	—	—
—	—	—	—



?

Learn by minimizing BIC

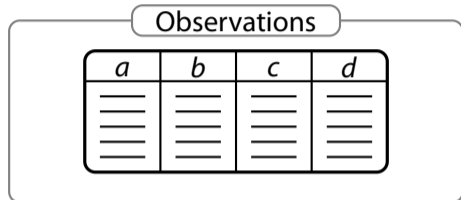
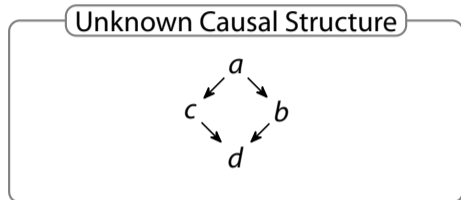


Assumptions

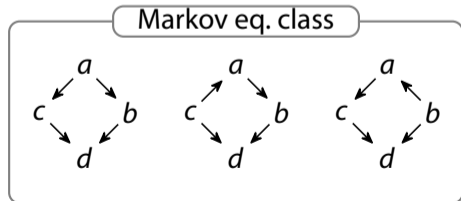
- Causal sufficiency
- Acyclicity
- Faithfulness
- Linear additive noise models

Causal Discovery

Sampling Process



Learn by minimizing BIC



Obstacles

- Identifiability/Markov equivalence
- Need for large sample sizes
- Large number of directed acyclic graphs
- NP-hardness of optimization problem

Combinatorial Hardness of Causal Discovery

Widespread Claim

The NP-hardness of learning directed acyclic graphs makes causal discovery intractable in practice.

Combinatorial Hardness of Causal Discovery

Widespread Claim

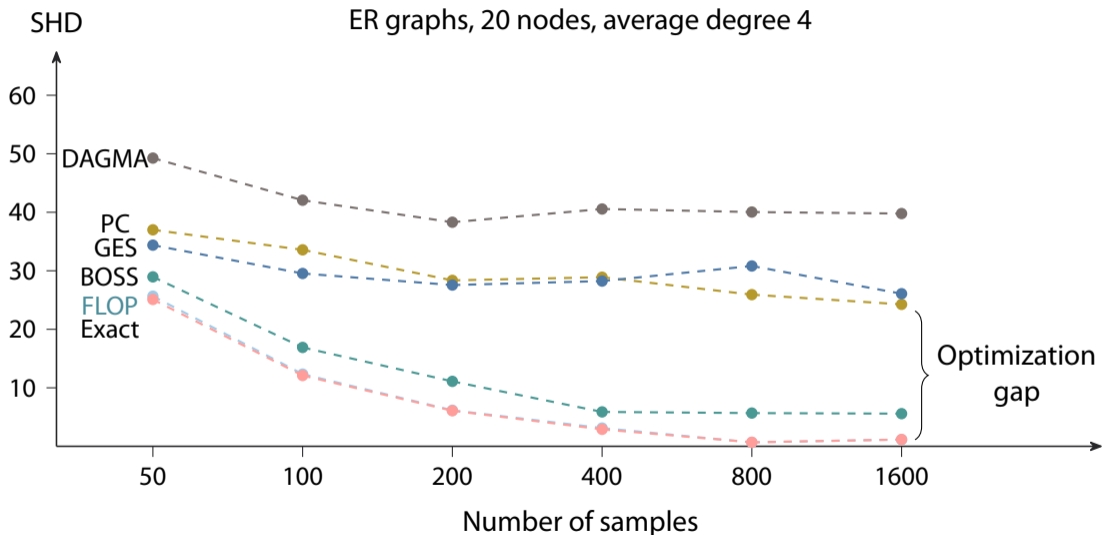
The NP-hardness of learning directed acyclic graphs makes causal discovery intractable in practice.

However:

- Real-world instances of NP-hard problems are often easily solved.
- There are lots of small-scale applications of causal discovery.
- NP-hardness constructions use latent variables, violating assumptions.
- In an oracle setting, sparse ground-truths can be recovered in polynomial-time.

So, what is the reason for poor performance on finite samples?

It's Not Only About the Sample Size



Embracing Discrete Search

Goal: Effective discrete search method for causal discovery of DAGs

Features:

- Scalable, efficient discrete local search
- Large neighborhoods to avoid local optima
- A metaheuristic to escape local optima

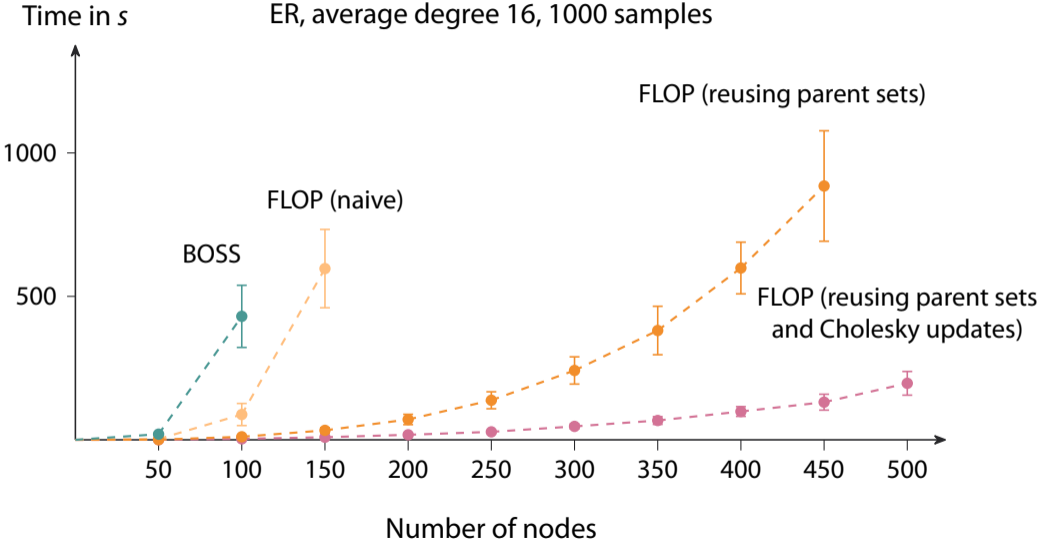
Building on the reinsertion-based order search from BOSS (Andrews et al., 2023), we develop **FLOP** (Fast Learning of Order and Parents).

Main Contributions

Components the FLOP algorithm:

- Parent selection with a non-greedy, warm-started grow-shrink routine
- Iterative scoring based on Cholesky updates
- Initial-order heuristic that addresses finite-sample issues of grow-shrink on a restricted variable set
- Iterated local search (ILS) to escape local optima

Runtime Improvements

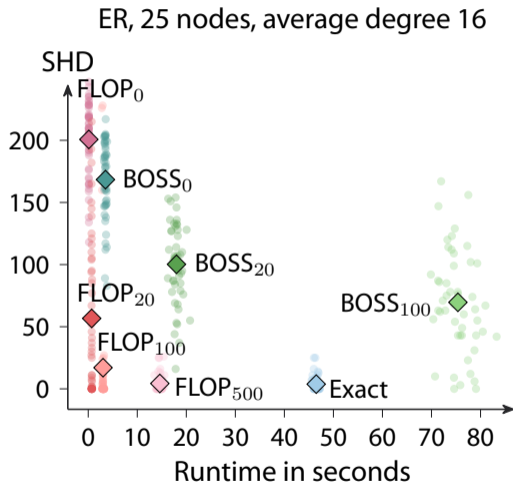
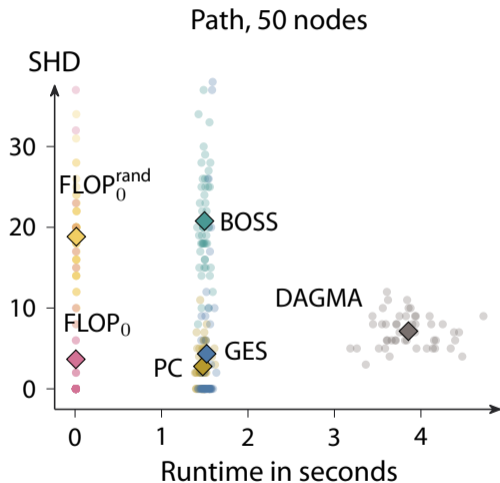


Main Contributions

Components the FLOP algorithm:

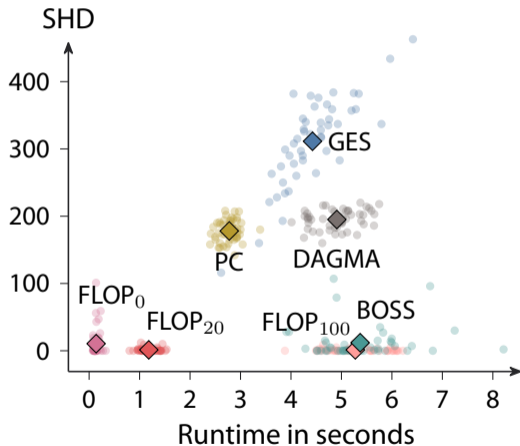
- Parent selection with a non-greedy, warm-started grow-shrink routine
- Iterative scoring based on Cholesky updates
- Initial-order heuristic that addresses finite-sample issues of grow-shrink on a restricted variable set
- Iterated local search (ILS) to escape local optima

Improvements on Paths and Dense Graphs

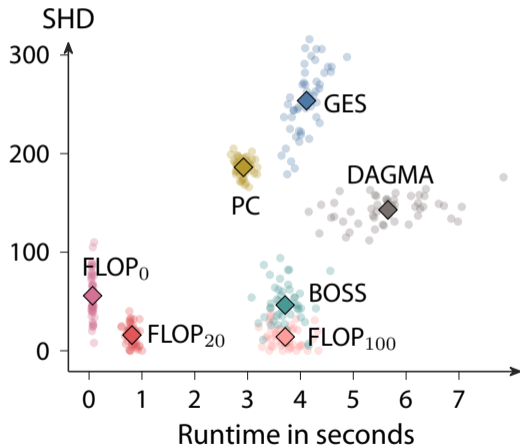


Standard Benchmarks

ER, 50 nodes, average degree 8



SF, 50 nodes, density parameter 4



Summary

- FLOP is largely effective at BIC optimization.
- On standard benchmark settings (ER graphs, linear ANM with Gaussian noise), this translates to high accuracy.
- There are cases (e.g., when assumptions are violated) where successful optimization of the linear BIC does not yield graphs close to the ground-truth.
- Thus, practical challenges of causal discovery remain:
 - Unobserved variables
 - Feedback loops
 - Identifiability
 - Faithfulness violations
 - Need for causal abstractions
 - Unknown ground-truth

