Two Optimal Strategies for Active Learning of Causal Models from Interventions

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

SfS Seminar for Statistics

Alain Hauser and Peter Bühlmann
Seminar for Statistics, ETH Zürich, http://stat.ethz.ch/~alhauser

Abstract

A causal DAG is in general not fully identifiable. Interventions generally improve identifiability; however, the gain of an intervention strongly depends on the intervened variables. We present active learning strategies calculating optimal interventions for two different learning goals:

1. maximizing the number of orientable edges with single-vertex interventions;
2. minimizing the number of interventions (at arbitrarily many variables) to guarantee full identifiability.

We compare our two active learning approaches to random interventions in a simulation study.

Causal Models, Interventions

Causal model: pair (D, f), D: DAG on vertex set $[p] := \{1, \ldots, p\}$, f: p-variate probability density with **Markov property** of D.

Example: D in Fig. 1 encodes the Markov property (i.e., factorization)

$$f(x) = f(x_1) f(x_2|x_1) f(x_3|x_1) f(x_4|x_2, x_3) \ .$$

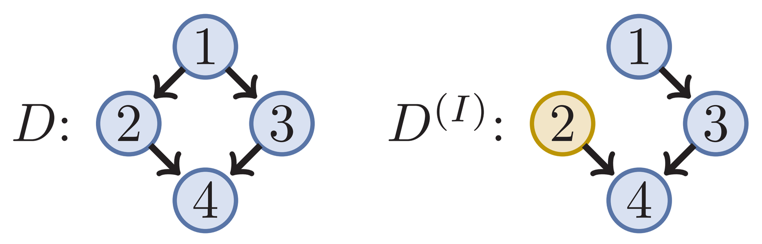


Fig. 1: DAG D and corresponding intervention DAG $D^{(I)}$ for the intervention target $I = \{2\}$.

Intervention at some **target** $I \subset [p]$: experiment which forces the variables X_I to values *independent* of the original causal parents.

Intervention DAG $D^{(I)}$: DAG D without all arrows pointing to a vertex in I (cf. Fig. 1). Encodes the Markov property of a causal model under an intervention at I.

(Interventional) Markov Equivalence

We consider a set of intervention experiments consisting of interventions at different targets summarized as a **family of targets** $\mathcal{I} \subset \mathcal{P}([p])$. Two DAGs D_1 and D_2 are **$\mathcal{I}$-Markov equivalent** (i.e. statistically indistinguishable under data from interventions at the targets in $\mathcal{I}$) if and only if (Hauser and Bühlmann, 2012)

(i) D_1 and D_2 have the same skeleton and the same v-structures, and

(ii) $D_1^{(I)}$ and $D_2^{(I)}$ have the same skeleton for all $I \in \mathcal{I}$.

$\mathcal{I}$-essential graph $\mathcal{E}_\mathcal{I}(D)$: partially directed graph representing the $\mathcal{I}$-Markov equivalence class of the DAG D:

- a **directed** edge represents an arrow with *common* (i.e., **identifiable**) orientation in all representatives;
- an **undirected** edge represents an arrow with *different* (i.e., **unidentifiable**) orientations in the representatives.

An essential graph G is a **chain graph** with **chordal** chain components (Hauser and Bühlmann, 2012); the **set** of its **representatives** is denoted by $\mathbf{D}(G)$.

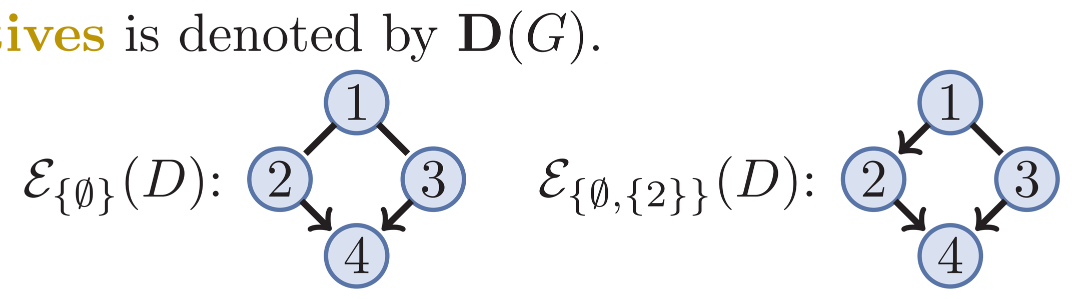


Fig. 2: Essential graphs of D (Fig. 1) for different families of targets.

Clique number $\omega(H)$: the size of the largest (undirected) clique in a graph H.

AL 1: Single-Vertex Targets

Problem: given an $\mathcal{I}$-essential graph G, find the single-vertex intervention target v such that the number of edges orientable after intervening at v is maximized.

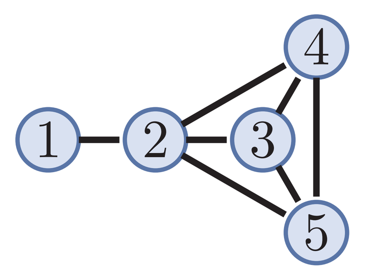
Formally: find

$$v := \arg\min_{v' \in [p]} \max_{D \in \mathbf{D}(G)} \xi\left(\mathcal{E}_{\mathcal{I} \cup \{\{v'\}\}}(D)\right) \ ;$$

$\xi(H)$: number of undirected edges in a graph H.

Solution: local algorithm OptSingle that considers the neighborhood of potential intervention targets v only, without enumerating all DAGs in $\mathbf{D}(G)$. OptSingle has in worst case exponential complexity (depending on $\omega(G)$).

Examples



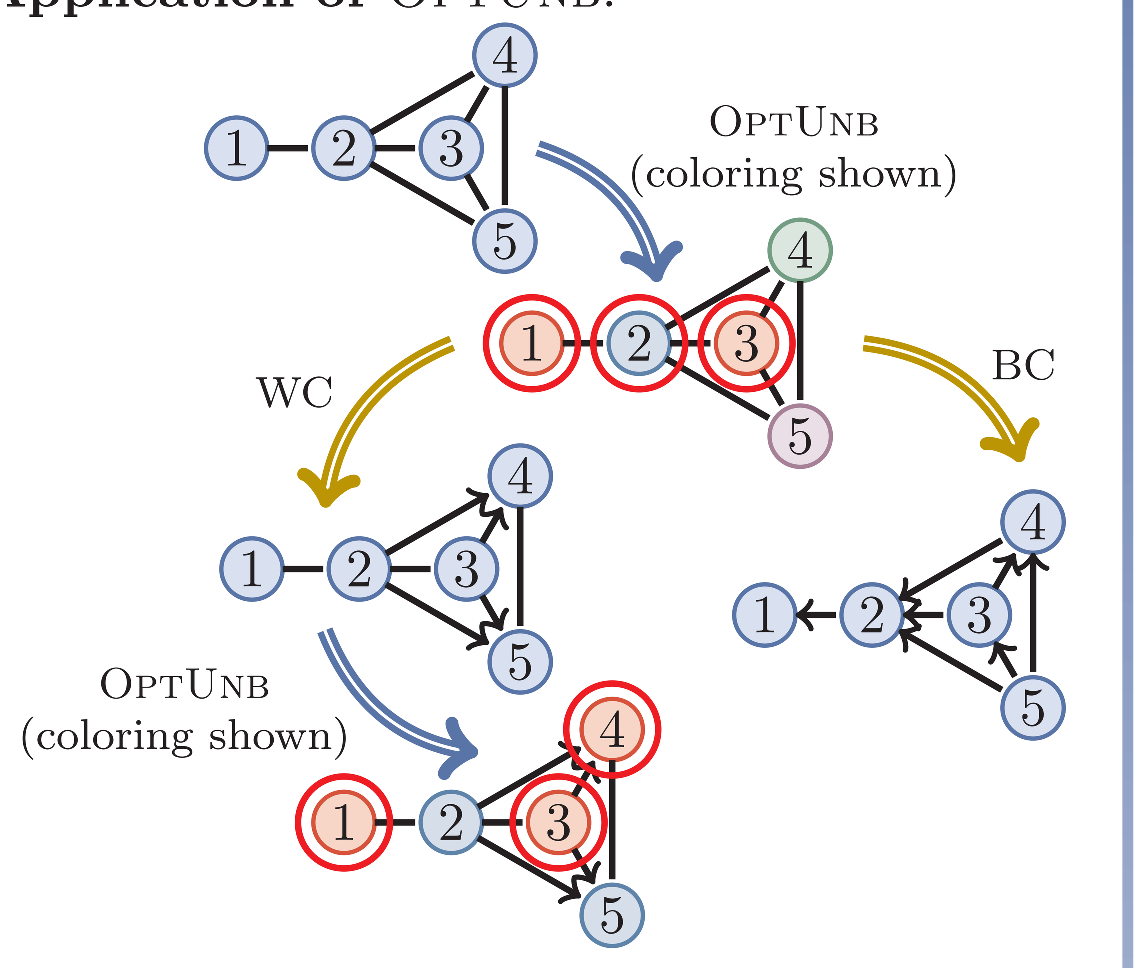
Left: (observational) essential graph

⇒: AL algorithm proposing intervention target ○

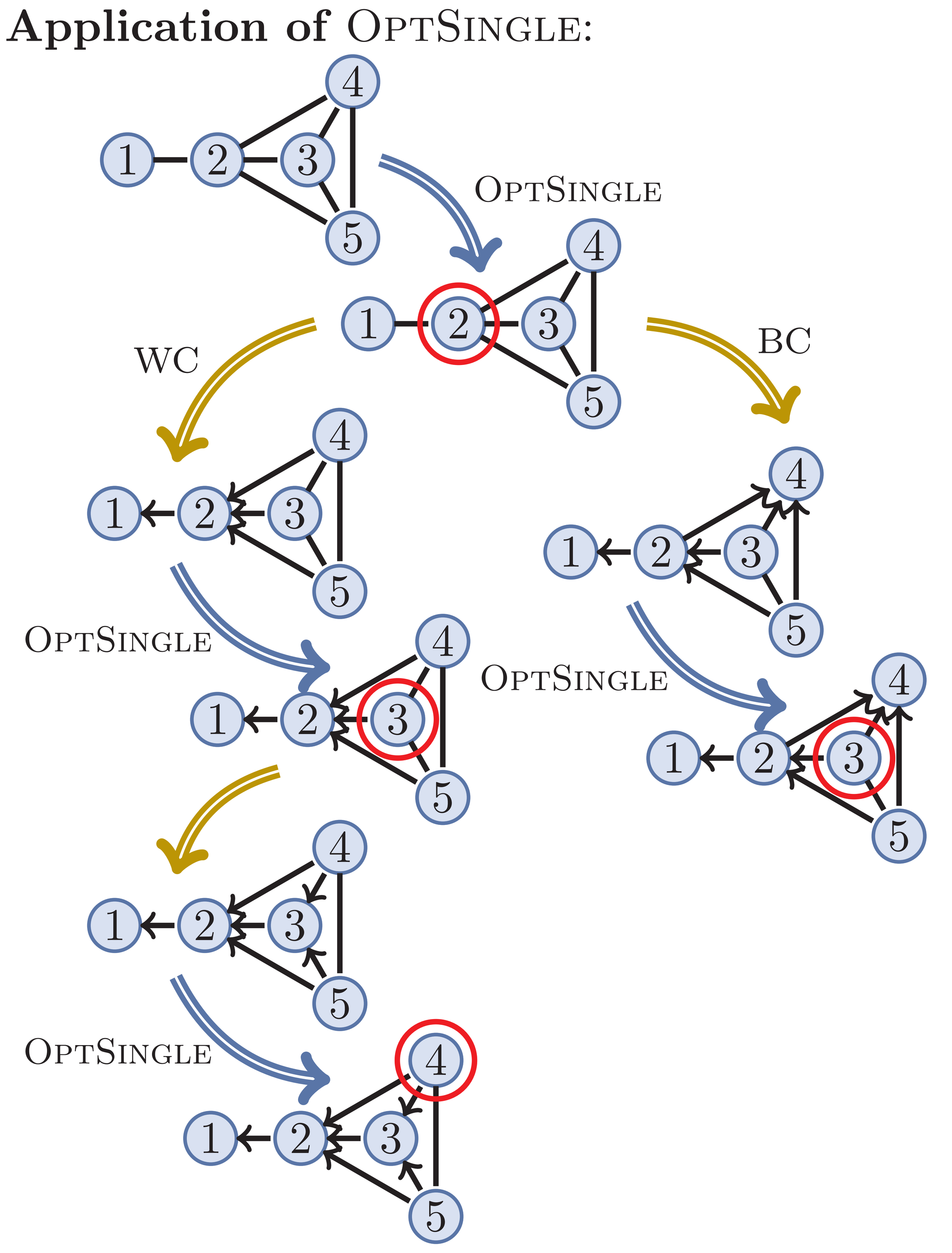
⇒: intervention at ○

WC: worst case; BC: best case

Application of OptUnb:



Application of OptSingle:



AL 2: Unbounded Targets

Problem: given an $\mathcal{I}$-essential graph G, find a minimum set of intervention targets $I_1, \ldots, I_k$ such that every DAG $D \in \mathbf{D}(G)$ is fully identifiable under $\mathcal{I}' := \mathcal{I} \cup \{I_1, \ldots, I_k\}$.

Solution: algorithm OptUnb that yields an intervention target I which **maximally reduces the clique size** of the essential graph after intervening. Iterative application of OptUnb yields a solution to the problem above; theoretical justification:

Prop. 1 *There is a target $I \subset [p]$ such that for every $D \in \mathbf{D}(G)$, $\omega(\mathcal{E}_{\mathcal{I} \cup \{I\}}(D)) \leq \lceil \omega(G)/2 \rceil$. This bound is sharp: for each $I \subset [p]$ there exists a DAG $D \in \mathbf{D}(G)$ such that $\omega(\mathcal{E}_{\mathcal{I} \cup \{I\}}(D)) \geq \lceil \omega(G)/2 \rceil$.*

This proves the conjecture of Eberhardt (2008):

Cor. 1 *$k = \lceil \log_2(\omega(G)) \rceil$ subsequent interventions are sufficient and in the worst case necessary for fully identifying every DAG in $\mathbf{D}(G)$.*

OptUnb is based on LexBFS (Rose, 1970) and greedy coloring; it has complexity $O(p + |E|)$ (E: edge set of G) by exploiting the chordality of the chain components of G.

Simulations

Aim: evaluation of OptUnb and OptSingle on randomly drawn causal models with $p \in \{10, 20, 30, 40\}$

Quality measure: **survival time**, i.e. number of steps needed for full identifiability; measured either in number of intervention targets (T) or total number of intervened variables (V).

Competing algorithms:

- Rand: purely random choice of single-vertex interventions
- RandAdv: random choice of single-vertex interventions among vertices with at least one incident undirected edge

Results:

- Rand clearly beaten by other algorithms
- OptUnb best algorithm in terms of intervention targets (T), OptSingle best algorithm in terms of intervened vertices (V)

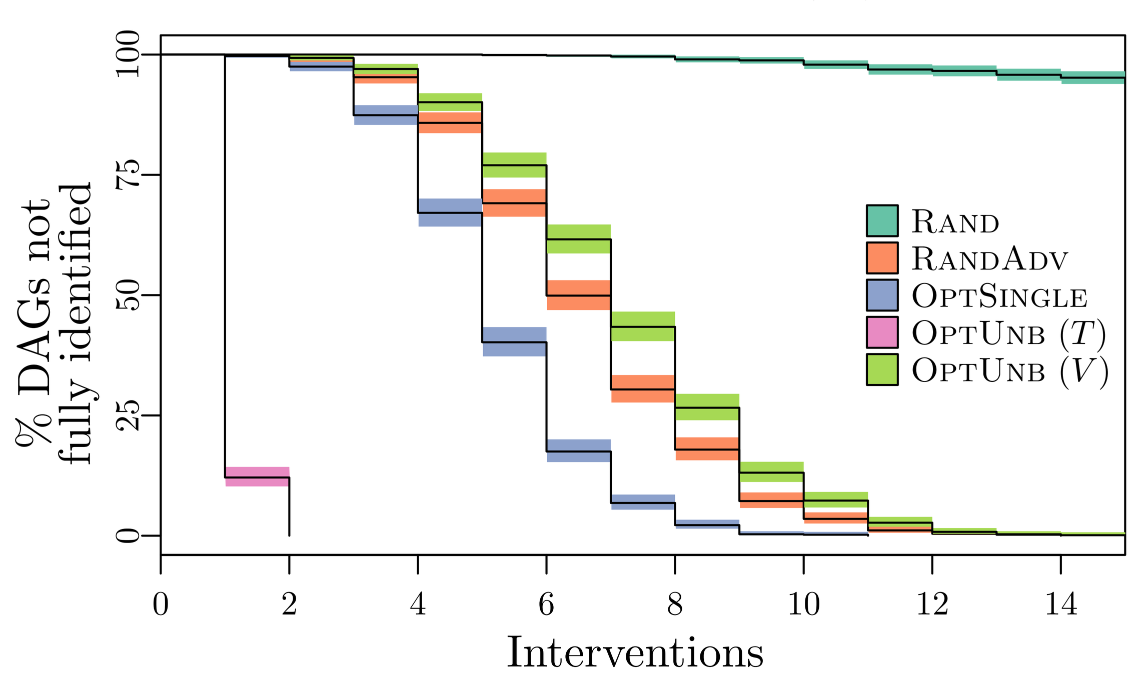


Fig. 3: Number of intervention steps needed for full identifiability of DAGs with $p = 40$, measured in targets (T) or intervened variables (V). Thin lines: Kaplan-Meier estimates (Kaplan and Meier, 1958); colored bands: 95% confidence region.

References

F. Eberhardt. Almost optimal intervention sets for causal discovery. In *UAI*, pages 161–168, 2008.

A. Hauser and P. Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *JMLR*, 13:2409–2464, 2012.

E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *JASA*, pages 457–481, 1958.

D.J. Rose. Triangulated graphs and the elimination process. *Journal of Mathematical Analysis and Applications*, 32(3):597–609, 1970.