Two Optimal Strategies for Active Learning of Causal Models from Interventions

Alain Hauser and Peter Bühlmann
ETH Zürich, Switzerland
{hauser, buhlmann}@stat.math.ethz.ch

Abstract

From observational data alone, a causal DAG is in general only identifiable up to Markov equivalence. Interventional data generally improves identifiability; however, the gain of an intervention strongly depends on the intervention target, i.e., the intervened variables. We present active learning strategies calculating optimal interventions for two different learning goals. The first one is a greedy approach using single-vertex interventions that maximizes the number of edges that can be oriented after each intervention. The second one yields in polynomial time a minimum set of targets of arbitrary size that guarantees full identifiability. This second approach proves a conjecture of Eberhardt (2008) indicating the number of unbounded intervention targets which is sufficient and in the worst case necessary for full identifiability. We compare our two active learning approaches to random interventions in a simulation study.

1 Introduction

Causal relationships between random variables are usually modeled by directed acyclic graphs (DAGs), where an arrow between two random variables, $X \longrightarrow Y$, reveals the former (X) as a *direct* cause of the latter (Y). From observational data alone (i.e. *passively* observed data from the undisturbed system), directed graphical models are only identifiable up to Markov equivalence, and arrow directions (which are crucial for the causal interpretation) are in general not identifiable. Without the assumption of specific functional model classes and error distributions (Peters et al., 2011), the only way to improve identifiability is to use interventional data for estimation, i.e. data produced under a perturbation of the system in which one or several random variables are forced to specific values, irrespective of the original causal parents.

The investigation of observational Markov equivalence classes has a long tradition in the literature (Verma and Pearl, 1990; Andersson et al., 1997; Spirtes et al., 2000). In a recent paper, Hauser and Bühlmann (2012) presented a graph-theoretic characterization of interventional Markov equivalence classes for a given set of interventions (possibly affecting several variables simultaneously). In this paper, we present two active learning strategies for finding valuable interventions: one that greedily optimizes the number of orientable edges with single-vertex interventions, and one that minimizes the number of interventions at arbitrarily many vertices to attain full identifiability.

Several approaches for actively learning causal models have been proposed during the last decade. Our method for finding intervention targets of *unbounded size* is closely related to the approach of Eberhardt (2008). In contrast to his procedure, our algorithm has a polynomial time complexity; furthermore, we prove his conjecture on the number of intervention experiments sufficient and in the worst case necessary for fully identifying a causal model. He and Geng (2008) presented a method to find a (nearly) minimal set of single-vertex interventions which guarantee the orientability of all undirected edges of an observational Markov equivalence class. In contrast to their approach, we proceed in a greedy way which

results in a smaller or at most equal number of single-vertex interventions to be performed. Tong and Koller (2001) finally proposed a Bayesian active learning strategy that minimizes an expected loss, in contrast to our frequentist methods.

This paper is organized as follows: in Sect. 2, we specify our notation of causal models and formalize our learning goals. In Sect. 3, we summarize graph-theoretic background material upon which our active learning algorithms, presented in Sect. 4, are based. In Sect. 5, we evaluate our algorithms in a simulation study.

2 Model

2.1 Causal Calculus

We consider a causal model on p random variables $(X_1, \dots, X_p)$ described by a DAG D. Formally, a causal model is a pair (D, f), where D is a DAG on the vertex set $[p] := \{1, \dots, p\}$ which encodes the **Markov property** of the (observational) density f: $f(x) = \prod_{i=1}^{p} f(x_i \mid x_{\mathrm{pa}_D(i)})$; $\mathrm{pa}_D(i)$ denotes the parent set of vertex i (see also Sect. 3). Unless stated otherwise, all graphs in this paper are assumed to have the vertex set $[p]$.

Beside the conditional independence relations of the observational density implied by the Markov property, a causal model also makes statements about effects of **interventions**. We consider **stochastic interventions** (Korb et al., 2004) modeling the effect of setting or forcing one or several random variables $X_I := (X_i)_{i \in I}$, where $I \subset [p]$ is called the **intervention target**, to the value of *independent* random variables U_I. Extending the do() operator (Pearl, 1995) to stochastic interventions, we denote the **interventional density** of X under such an intervention by

$$f(x|\,\mathrm{do}_D(X_I = U_I)) := \prod_{i \notin I} f(x_i | x_{\mathrm{pa}_D(i)}) \prod_{i \in I} \tilde{f}(x_i),$$

where $\tilde{f}$ is the density of U_I on $\mathcal{X}_I$. By denoting with $I = \emptyset$ and using the convention $f(x|\,\mathrm{do}(X_\emptyset = U_\emptyset)) = f(x)$, we also encompass the observational case as an intervention target. The interventional density $f(x|\,\mathrm{do}_D(X_I = U_I))$ has the Markov property of the **intervention graph** $D^{(I)}$, the DAG that we get from D by removing all arrows pointing to vertices in I.

We consider experiments based on datasets originating from *multiple* interventions. The **family of targets** $\mathcal{I} \subset \mathcal{P}([p])$, where $\mathcal{P}([p])$ denotes the power set of $[p]$, lists all (distinct) intervention targets used in an experiment. A family of targets $\mathcal{I} = \{\emptyset, \{3\}, \{1, 4\}\}$ e.g. characterizes an experiment in which observational data as well as data originating from an intervention at X_3 and data originating from a (simultaneous) intervention at X_1 and X_4 are measured. In the whole paper, $\mathcal{I}$ always stands for a family of targets with the property that for each vertex $a \in [p]$, there is some target $I \in \mathcal{I}$ in which a is *not* intervened ($a \notin I$). This is e.g. the case when observational data is available ($\emptyset \in \mathcal{I}$). Two DAGs D_1 and D_2 are called **$\mathcal{I}$-Markov equivalent** ($D_1 \sim_{\mathcal{I}} D_2$) if they are statistically indistinguishable under an experiment consisting of interventions at the targets in $\mathcal{I}$; we refer to Hauser and Bühlmann (2012) for a more formal treatment.

Theorem 1 (Hauser and Bühlmann (2012))**.** *Two DAGs D_1 and D_2 are $\mathcal{I}$-Markov equivalent if and only if*
(i) D_1 and D_2 have the same skeleton and the same v-structures (that is, induced subgraphs of the form $a \longrightarrow b \longleftarrow c$), and
(ii) $D_1^{(I)}$ and $D_2^{(I)}$ have the same skeleton for all $I \in \mathcal{I}$.

An $\mathcal{I}$-Markov equivalence class of a DAG D is uniquely represented by its **$\mathcal{I}$-essential graph** $\mathcal{E}_{\mathcal{I}}(D)$ (Hauser and Bühlmann, 2012). This partially directed graph has the same skeleton as D; a directed edge in $\mathcal{E}_{\mathcal{I}}(D)$ represents **$\mathcal{I}$-essential** arrows, i.e. arrows that have the same orientation in all DAGs of the equivalence class; an undirected edge represents arrows that have different orientations in different DAGs of the equivalence class. The concept of $\mathcal{I}$-essential graphs generalizes the one of CPDAGs which is well-known in the observational case (Spirtes et al., 2000; Andersson et al., 1997). We denote the $\mathcal{I}$-Markov equivalence class corresponding to an $\mathcal{I}$-essential graph G by $\mathbf{D}(G)$.

2.2 Active Learning

Assume G is an $\mathcal{I}$-essential graph estimated from interventional data produced under the different interventions in $\mathcal{I}$. We consider two different greedy active learning approaches. In one step, the first one computes a single-vertex intervention that maximizes the number of orientable edges, while the second one computes an intervention target of arbitrary size that maximally reduces the clique number, i.e. the size of the largest clique of undirected edges (see Sect. 3). The motivation for the first approach is the attempt to quickly improve the identifiability of causal models with interventions at few variables; the motivation for the second approach is the conjecture of Eberhardt (2008) (which we prove in Cor. 2) stating that maximally reducing the clique number after each intervention yields full identifiability of causal models with a minimal number of interventions.

Formally, our two algorithms yield a solution to the following problems. The first one, called OptSingle, computes a vertex

$$v = \arg\min_{v' \in [p]} \max_{D \in \mathbf{D}(G)} \xi\left(\mathcal{E}_{\mathcal{I} \cup \{\{v'\}\}}(D)\right) \ , \qquad (1)$$

where $\xi(H)$ denotes the number of undirected edges in a graph H. The second algorithm, called OptUnb, computes a set

$$I = \arg\min_{I' \subset [p]} \max_{D \in \mathbf{D}(G)} \omega\left(\mathcal{E}_{\mathcal{I} \cup \{I'\}}(D)\right) \ , \qquad (2)$$

where $\omega(H)$ denotes the clique number of H (see also Sect. 3). The key ingredients for the efficiency of OptSingle (Alg. 2) and OptUnb (Alg. 3) are implementations that minimize the objective functions of Eq. (1) and (2), resp., without enumerating all DAGs in the equivalence class represented by G. Graph theoretic results upon which our implementations are based are summarized in the next section.

3 Graph Theoretic Background

A **graph** is a pair $G = (V, E)$, where V is a set of vertices and $E \subset (V \times V) \setminus \{(a, a) | a \in V\}$ is a set of edges. We always assume $V = [p] := \{1, 2, \ldots, p\}$ and let the vertices of a graph represent the p random variables $X_1, \ldots, X_p$.

An edge $(a, b) \in E$ with $(b, a) \in E$ is called **undirected** and denoted by $a - b$, whereas an edge $(a, b) \in E$ with $(b, a) \notin E$ is called **directed** and denoted by $a \longrightarrow b$. G is called directed if all its edges are directed (or undirected, resp.). A **cycle** of length $k \geq 2$ is a sequence of k distinct vertices of the form $(a_0, a_1, \ldots, a_k = a_0)$ such that $(a_{i-1}, a_i) \in E$ for $i \in \{1, \ldots, k\}$; the cycle is **directed** if at least one edge is directed.

For a subset $A \subset V$ of the vertices of G, the **induced subgraph** on A is $G[A] := (A, E[A])$, where $E[A] := E \cap (A \times A)$. A **v-structure** is an induced subgraph of G of the form $a \longrightarrow b \longleftarrow c$. The **skeleton** of a graph G is the undirected graph $G^u := (V, E^u)$, $E^u := E \cup \{(a, b) \mid (b, a) \in E\}$. The **parents** of a vertex $a \in V$ are the vertices $\mathrm{pa}_G(a) := \{b \in V \mid b \longrightarrow a \in G\}$, its **neighbors** are the vertices $\mathrm{ne}_G(a) := \{b \in V \mid a - b \in G\}$; the **degree** of a is defined as $\deg(a) := |\{b \in V \mid (a, b) \in E \vee (b, a) \in E\}|$, the maximum degree of G is $\Delta(G) := \max_{a \in V} \deg(a)$.

An undirected graph $G = (V, E)$ is **complete** if all pairs of vertices are neighbors. A **clique** is a subset of vertices $C \subset V$ such that $G[C]$ is complete. The **clique number** $\omega(G)$ of G is the size of the largest clique in G. G is **chordal** if every cycle of length $k \geq 4$ contains a **chord**, i.e. two nonconsecutive adjacent vertices.

A **directed acyclic graph** or **DAG** is a directed graph without cycles. A partially directed graph $G = (V, E)$ is a **chain graph** if it contains no *directed* cycle; undirected graphs and DAGs are special cases of chain graphs. Let G' be the undirected graph we get by removing all directed edges from a chain graph G. The **chain component** $T_G(a)$ of a vertex a is the set of all vertices that are connected to a in G'. The set of all chain components of G is denoted by $\mathbf{T}(G)$; they form a partition of V. We extend the clique number to chain graphs G by the definition $\omega(G) := \max_{T \in \mathbf{T}(G)} \omega(G[T])$.

An **ordering** of a graph G, i.e. a permutation $\sigma : [p] \to [p]$, induces a total order on V by the definition $a \leq_\sigma b :\Leftrightarrow \sigma^{-1}(a) \leq \sigma^{-1}(b)$. An ordering $\sigma = (v_1, \ldots, v_p)$ is a **perfect elimination ordering** or **PEO** if for all i, $\mathrm{ne}_{G^u} \cap \{v_1, \ldots, v_{i-1}\}$ is a clique in G^u. A **topological ordering** or **TO** of a DAG D is an

Input : $G = ([p], E)$: undirected graph;
$\sigma = (v_1, \ldots, v_p)$: ordering of G.
Output: A proper coloring $c : [p] \to \mathbb{N}$
$c(v_1) \leftarrow 1$;
for $i = 2$ *to* p **do**
  $c(v_i) \leftarrow \min\{k \in \mathbb{N} \mid k \neq c(u)\ \forall\ u \in \{v_1, \ldots, v_{i-1}\} \cap \mathrm{ne}(v_i)\}$;
return c;

Algorithm 1: GreedyColoring(G, σ). Greedy algorithm that yields a proper coloring of G along an ordering σ.

ordering σ such that $a \leq_\sigma b$ for each arrow $a \longrightarrow b \in D$; we then say that D is **oriented according to** σ.

For the rest of this section, let $G = (V, E)$ be an undirected graph. We consider a variant of the breadth-first search (BFS) called **lexicographic BFS** or LexBFS (Rose, 1970) that takes an ordering $(v_1, \ldots, v_p)$ of V and the edge set E as input and that outputs an ordering $\sigma = \mathrm{LexBFS}((v_1, \ldots, v_p), E)$ listing the vertices of V in the visited order. If $\{v_1, \ldots, v_k\}$ is a clique, σ also starts with $v_1, \ldots, v_k$; we refer to Hauser and Bühlmann (2012) for details of such an implementation. For a set $A = \{a_1, \ldots, a_k\} \subset V$ and an additional vertex $v \in V \setminus A$, e.g., we use the notation $\mathrm{LexBFS}((A, v, \ldots), E)$ to denote a LexBFS-ordering produced from a start order of the form $(a_1, \ldots, a_k, v, \ldots)$, without specifying the orderings of A and $V \setminus (A \cup \{v\})$.

Proposition 1 (Rose et al. (1976))**.** *Let $G = (V, E)$ be an undirected chordal graph with a* LexBFS*-ordering σ. Then σ is also a PEO on G. By orienting the edges of G according to σ, we get a DAG without v-structures.*

Alg. 3 is strongly based on graph colorings. A k**-coloring** of G is a map $c : V \to [k]$; the coloring c is **proper** if $c(u) \neq c(v)$ for every edge $u - v \in G$. We say that G is k**-colorable** if it admits a proper k-coloring; the **chromatic number** $\chi(G)$ of G is the smallest integer k such that G is k-colorable. By greedily coloring the vertices of the graph (see Alg. 1), one gets a proper k-coloring with $k \leq \Delta(G) + 1$ in polynomial time (Chvátal, 1984).

For any undirected graph G, the bounds $\omega(G) \leq \chi(G) \leq \Delta(G) + 1$ hold. G is **perfect** if $\omega(H) = \chi(H)$ holds for every induced subgraph H of G. An ordering σ of G is **perfect** if for any induced subgraph H of G, greedy coloring along the ordering induced by σ yields an optimal coloring of H (i.e., a $\chi(H)$-coloring).

Proposition 2 (Chvátal (1984))**.** *An ordering σ of an undirected graph G is perfect if and only if G has no induced subgraph of the form $a - b - c - d$ with $a <_\sigma b$ and $d <_\sigma c$.*

It can easily be seen that a PEO fulfills the requirement of Prop. 2; hence we get, together with Prop. 1, the following corollary.

Corollary 1. *(i) A perfect elimination ordering on some graph G is perfect.*
(ii) Any chordal graph has a perfect ordering.

Proposition 3 (Chvátal (1984))**.** *A graph with a perfect ordering is perfect.*

4 Optimal Intervention Targets

An $\mathcal{I}$-essential graph is a chain graph with chordal chain components. Their edges are oriented according to a PEO in the DAGs of the corresponding equivalence class; edge orientations of different chain components do not influence (i.e., additionally restrict) each other (Hauser and Bühlmann (2012), Thm. 18 and Prop. 16). We can thus restrict our search for optimal intervention targets to single chain components. We can even treat each chain component as an observational essential graph, as the following lemma shows; we skip a formal proof which is rather simple, but technical.

Lemma 1. *Consider an $\mathcal{I}$-essential graph $\mathcal{E}_\mathcal{I}(D)$ of some DAG D, and let $T \in \mathbf{T}(\mathcal{E}_\mathcal{I}(D))$. Furthermore, let $I \subset [p]$, $I \notin \mathcal{I}$, be an (additional) intervention target. Then we have*

$$\mathcal{E}_{\mathcal{I} \cup \{I\}}(D)[T] = \mathcal{E}_{\{\emptyset, I \cap T\}}(D[T])\ .$$

4.1 Single-vertex Interventions

We start with the treatment of the first active learning approach mentioned in Sect. 2.2. By virtue of the following lemma, the maximum in Eq. (1) can be calculated without enumerating all representative DAGs. The lemma easily follows from Thm. 1; we skip the proof.

Lemma 2. *Let G be an $\mathcal{I}$-essential graph, and let $v \in [p]$. Assume D_1 and $D_2 \in \mathbf{D}(G)$ such that $\{a \in \mathrm{ne}_G(v) \mid a \longrightarrow v \in D_1\} = \{a \in \mathrm{ne}_G(v) \mid a \longrightarrow v \in D_2\} = C$. Then we have $D_1 \sim_{\mathcal{I}'} D_2$ under the family of targets $\mathcal{I}' = \mathcal{I} \cup \{\{v\}\}$.*

The next proposition states that every clique in $\mathrm{ne}_G(v)$ is an admissible set C in the sense of Lem. 2 and vice versa. Algorithmically, a DAG D as described in Prop. 4 can be constructed with LexBFS; this motivates Alg. 2 which yields a solution of Eq. (1).

Proposition 4 (Andersson et al. (1997))**.** *Let G be an undirected chordal graph, $a \in [p]$ and $C \subset \mathrm{ne}_G(a)$. There is a DAG $D \subset G$ with $D^u = G$ and $\{b \in \mathrm{ne}(a) \mid b \longrightarrow a \in D\} = C$ which is oriented according to a PEO if and only if C is a clique.*

4.2 Interventions at Targets of Arbitrary Size

We now proceed to the solution of Eq. (2). The following proposition, which was already conjectured by Eberhardt (2008), shows that the minimum in Eq. (2) only depends on the clique number of G:

$$\min_{I' \subset [p]} \max_{D \in \mathbf{D}(G)} \omega\left(\mathcal{E}_{\mathcal{I} \cup \{I'\}}(D)\right) = \lceil \omega(G)/2 \rceil .$$

Input : $G = ([p], E)$: $\mathcal{I}$-essential graph.
Output: An optimal intervention vertex $v \in [p]$, or $\emptyset$ if G only has directed edges.

```
v_opt ← 0; η_opt ← 0;
for v = 1 to p do
    η_min ← p^2;
    foreach clique C ⊂ ne_G(v) do
        σ ← LexBFS((C, v, ...)), E[T_G(v)]);
        D ← DAG with skeleton G, topological ordering σ;
        G' ← E_{∅,{v}}(D);
        η ← number of arrows in G';
        if η < η_min then η_min ← η;
    if p^2 > η_min > η_opt then
    (v_opt, η_opt) ← (v_opt, η_min);
if v_opt ≠ 0 then return v_opt;
else return ∅;
```

Algorithm 2: OptSingle(G): yields a solution of Eq. (1).

Proposition 5. *Let G be an undirected, connected, chordal graph on the vertex set $V = [p]$; such a graph is an observational essential graph.*
(i) There is an intervention target $I \subset V$ such that for every DAG $D \in \mathbf{D}(G)$, we have

$$\omega(\mathcal{E}_{\{\emptyset, I\}}(D)) \leq \lceil \omega(G)/2 \rceil .$$

(ii) For every intervention target $I \subset [p]$ there is a DAG $D \in \mathbf{D}(G)$ such that

$$\omega(\mathcal{E}_{\{\emptyset, I\}}(D)) \geq \lceil \omega(G)/2 \rceil .$$

Proof. (i) Since G is chordal, we have $\chi(G) = \omega(G)$ by Cor. 1(ii) and Prop. 3. Let $c : V \to [\omega(G)]$ be a proper coloring of G. Define $I := c^{-1}([h])$ for $h := \lceil \omega(G)/2 \rceil$. With an intervention at the target I, at most the edges of $G[I]$ and $G[V \setminus I]$ are unorientable for any causal structure $D \in \mathbf{D}(G)$ under the family of targets $\mathcal{I} := \{\emptyset, I\}$. Therefore the bound

$$\omega(\mathcal{E}_{\mathcal{I}}(D)) \leq \max\{\omega(G[I]), \omega(G[V \setminus I])\}$$

holds for every $D \in \mathbf{D}(G)$. It remains to show that both of these terms are bounded by h.

The induced subgraph $G[I]$ is also perfect, and $c|_I$ is a proper h-coloring of $G[I]$. Hence we have $\omega(G[I]) = \chi(G[I]) \leq h$. Analogously, $c|_{V \setminus I}$ is a proper $(\omega(G) - h)$-coloring of $G[V \setminus I]$, hence we have $\omega(G[V \setminus I]) = \chi(G[V \setminus I]) \leq \omega(G) - h \leq h$ by definition of h.
(ii) Let C be a maximum clique in G, and define $C \cap I =: \{v_1, \ldots, v_k\}$ and $C \setminus I =: \{v_{k+1}, \ldots, v_\omega\}$. The LexBFS-ordering $\sigma := \mathrm{LexBFS}((v_1, \ldots, v_\omega, \ldots), E)$ starts with the vertices $v_1, \ldots, v_\omega$. Set $\mathcal{I} := \{\emptyset, I\}$ and let $D \in \mathbf{D}(G)$ be oriented according to σ.

We claim that the arrows in $D[C \cap I]$ and in $D[C \setminus I]$ are not $\mathcal{I}$-essential in D. For $v_i, v_j \in C \cap I$ $(i < j)$, consider the ordering $\sigma' := \mathrm{LexBFS}((v_1, \ldots, v_j, \ldots, v_i, \ldots, v_\omega, \ldots), E)$, and $D' \in \mathbf{D}(G)$ which is obtained by orienting the edges of G according to σ'. We then have $D' \sim_{\mathcal{I}} D$:

- D and D' obviously have the same skeleton, and both have no v-structures.
- $D^{(I)}$ and $D'^{(I)}$ have the same skeleton because all arrows between a vertex $a \in I$ and another one $b \notin I$ point away from a.

Input : $G = ([p], E)$: essential graph.
Output: An optimal intervention target $I \subset [p]$.
$I \leftarrow \emptyset$;
foreach $T \in \mathbf{T}(G)$ **do**
  $\sigma \leftarrow \text{LexBFS}(T, E[T])$;
  $c \leftarrow \text{GreedyColoring}(G[T], \sigma)$;
  $\omega \leftarrow \max_{v \in [p]} c(v)$; $h \leftarrow \lceil \omega/2 \rceil$;
  $I \leftarrow I \cup c^{-1}([h])$;
return I;

Algorithm 3: OptUnb(G): yields a solution of Eq. (2); time complexity is $O(p + |E|)$.

For $v_i, v_j \in C \setminus I$, the argument is analogous, which proves the claim.

$\mathcal{E}_{\mathcal{I}}(D)$ contains the cliques $C \cap I$ and $C \setminus I$ of size k and $\omega(G) - k$ though. The fact that $\max\{k, \omega(G) - k\} \geq \lceil \omega(G)/2 \rceil$ completes the proof. □

The constructive proof shows that a minimizer I of Eq. (2) can be generated by means of an optimal coloring which we can get by greedy coloring along a LexBFS-ordering (see Prop. 1 and Cor. 1); this justifies Alg. 3.

Since an $\mathcal{I}$-essential graph has only one representative DAG if and only if its clique number is 1, a direct consequence of Prop. 5 is a (sharp) upper bound on the number of interventions necessary to fully identify a causal model, as it was conjectured by Eberhardt (2008).

Corollary 2. *Let G be the an $\mathcal{I}$-essential graph. There is a set of $k = \lceil \log_2(\omega(G)) \rceil$ intervention targets $I_1, \ldots, I_k$ which are sufficient and in the worst case necessary to make the causal structure fully identifiable:*

$$\mathcal{E}_{\mathcal{I} \cup \{I_1, \ldots, I_k\}}(D) = D \; \forall \; D \in \mathbf{D}(G).$$

The intervention targets $I_1, \ldots, I_k$ of Cor. 2 can be constructed by iteratively running Alg. 3 on $G = \mathcal{E}_{\mathcal{I}}(D)$, $\mathcal{E}_{\mathcal{I} \cup \{I_1\}}(D)$, $\mathcal{E}_{\mathcal{I} \cup \{I_1, I_2\}}(D)$ etc. However, they could also be constructed at once by a modification of Alg. 3. Let $c : [p] \to [\omega(G)]$ be a function such that for each chain component $T \in \mathbf{T}(G)$, $c|_T$ is a proper coloring of $G[T]$. By defining I_j as the set of all vertices whose color has a 1 in the j^{th} position of its binary representation, we make sure that for every pair of neighboring vertices u and v, there is at least one j such that $|\{u, v\} \cap I_j| = 1$; hence the edge between u and v is orientable in $\mathcal{E}_{\mathcal{I} \cup \{I_1, \ldots, I_k\}}(D)$. Since the binary representation of $\omega(G)$, the largest color in c, has length $k = \lceil \log_2(\omega(G)) \rceil$, this procedure creates a set of k intervention targets that fulfill the requirements of Cor. 2.

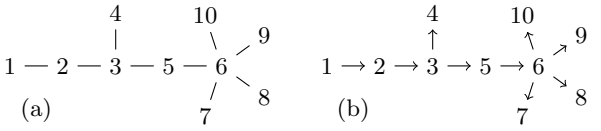

Figure 1: Observational essential graph (a) and a representative (b).

The problem of finding intervention targets to fully identify a causal model is related to the problem of finding separating systems of the chain components of essential graphs (Eberhardt, 2007). A **separating system** of an undirected graph $G = (V, E)$ is a subset $\mathcal{F}$ of the powerset of V such that for each edge a—$b \in G$, there is a set $F \in \mathcal{F}$ with $|F \cap \{a, b\}| = 1$. Cai (1984) has shown that the minimum separating system of a graph G has cardinality $\lceil \log_2(\chi(G)) \rceil$; this also proves Cor. 2. The proof of Cai (1984) uses arguments similar to ours given in the paragraph above for the non-iterative determination of the targets $I_1, \ldots, I_k$ of Cor. 2.

4.3 Discussion

LexBFS and GreedyColoring have a time complexity of $O(p + |E|)$ when executed on a graph $G = ([p], E)$. Thus, OptUnb (Alg. 3) also has a linear complexity.[1] The time complexity of OptSingle (Alg. 2) on the other hand depends on the size of the largest clique in the $\mathcal{I}$-essential graph G. By restricting OptSingle to $\mathcal{I}$-essential graphs with a bounded degree, its complexity is polynomial in p; otherwise, it is in the worst case exponential.

We emphasize that our two active learning algorithms do *not* optimize the same objective; OptUnb does *not* guarantee maximal identifiability after each intervention, and OptSingle

[1] In contrast, finding a minimum separating set on *non-chordal* graphs is NP-complete (Cai, 1984).

does *not* guarantee a minimal number of single-vertex interventions to full identifiability. Consider e.g. the (observational) essential graph in Fig. 1(a). It is not hard to see that all its representatives are fully identifiable after at most two single-vertex interventions: the first intervention should be performed at vertex 3, the second one either at vertex 2 or 6. If the DAG in Fig. 1(b) represents the true causal model, however, OptSingle will need three steps to full identifiability; it will iteratively propose interventions at targets 6, 3 and 2.

In general, OptUnb does not yield an intervention target of minimal size. With two straightforward improvements, we could reduce the number of intervened vertices: first, we could take $h \leftarrow \lfloor \omega/2 \rfloor$ instead of $h \leftarrow \lceil \omega/2 \rceil$ in Alg. 3; the proof of Prop. 5 is also valid with this choice. Second, we could permute the colors produced by the greedy coloring such that $|c^{-1}(\{1\})| \leq |c^{-1}(\{2\})| \leq \ldots$. However, since an optimal coloring of a graph is not unique, not even up to permutation of colors, these heuristic improvements would still not guarantee a *minimal* intervention target with the properties required in Prop. 5.

5 Experimental evaluation

We evaluated Alg. 2 and 3 in a simulation study on 4000 randomly generated causal models with vertex numbers $p \in \{10, 20, 30, 40\}$.

5.1 Methods

We compared four active learning approaches: our two algorithms OptSingle and OptUnb, a purely random proposition of single-vertex interventions (denoted by Rand), and a slightly advanced random approach that randomly chooses any vertex which has at least one incident undirected edge (denoted by RandAdv). To measure the quality of the proposed interventions, we evaluated the algorithms together with an "oracle estimator", that is, an algorithm that yields the true $\mathcal{I}$-essential graph of some DAG. This corresponds to model estimation in the limit of infinite sample sizes.

For each vertex number $p = \{10, 20, 30, 40\}$, we randomly generated 1000 DAGs with a binomial distribution of vertex degrees, having an expected degree of 3. Starting from the observational essential graph, we iteratively included the intervention targets proposed by the active learning algorithms. We defined the "survival time" of a DAG as the number of active learning steps needed for full identifiability, measured in intervention targets (T) or intervened variables (V). If a DAG was fully identifiable e.g. under the family $\mathcal{I} = \{\emptyset, \{1, 4\}, \{3\}\}$, we counted $T = 2$ (non-empty) targets and $V = 3$ variables. For each vertex number and algorithm, we estimated the "survival function", i.e. the probability $S_T(t) := P[T > t]$ or $S_V(v) := P[V > v]$, resp., with a Kaplan-Meier estimator (Kaplan and Meier, 1958).

5.2 Results

Fig. 2 shows the estimated survival functions of the active learning algorithms. Rand was clearly beaten by all competitors, and OptUnb dominated all other strategies in terms of intervention targets. However, if we measure the number of intervened vertices, OptUnb was even slightly worse than RandAdv. OptSingle gave a significant improvement over RandAdv; however, the step from Rand to RandAdv is much larger than from Rand to OptSingle.

When essential graphs are not given by an oracle, but estimated from finite samples with e.g. Greedy Interventional Equivalence Search (Hauser and Bühlmann, 2012), the convergence to the true model is slower due to estimation errors. Performance differences e.g. between RandAdv and OptSingle vanish for small sample sizes (data not shown).

6 Conclusion

We developed two algorithms which propose optimal intervention targets: one that finds the single-vertex intervention which maximally increases the number of orientable edges (called OptSingle), and one that maximally reduces the clique number of the non-orientable edges with an intervention at arbitrarily many variables (called OptUnb). We proved a conjecture of Eberhardt (2008) concerning the number of

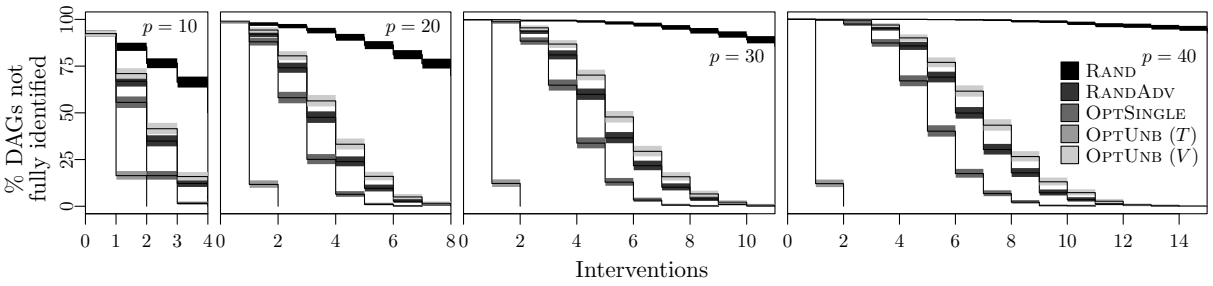

Figure 2: Number of intervention steps needed for full identifiability of DAGs, measured in targets (T) or intervened variables (V); for algorithms proposing only single-vertex interventions, both numbers are the same. Thin lines: Kaplan-Meier estimates; colored bands: 95% confidence region.

interventions sufficient and in the worst case necessary for fully identifying a causal model by showing that the OptUnb yields, when applied iteratively, a *minimum* set of intervention targets that guarantee full identifiability.

In a simulation study, we showed that both algorithms lead significantly faster to full identifiability than randomly chosen interventions. If we count the total number of intervened variables, however, OptUnb performed slightly worse than a random approach. This illustrates the fact that sequentially intervening single variables yields in general more identifiability that intervening those variables simultaneously.

Acknowledgments

We thank Jonas Peters, Frederick Eberhardt and the anonymous reviewers for valuable comments on the manuscript.

References

S.A. Andersson, D. Madigan, and M.D. Perlman. 1997. A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Stat.*, 25(2):505–541.

M.-C. Cai. 1984. On separating systems of graphs. *Disc. Math.*, 49:15–20.

V. Chvátal. 1984. Perfectly ordered graphs. *Ann. Disc. Math.*, 21:63–68.

F. Eberhardt. 2007. *Causation and Intervention.* Ph.D. thesis, Carnegie Mellon University.

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

A. Hauser and P. Bühlmann. 2012. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *To appear in JMLR; arxiv preprint arXiv:1104.2808.*

Y.-B. He and Z. Geng. 2008. Active learning of causal networks with intervention experiments and optimal designs. *JMLR*, 9:2523–2547.

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

K.B. Korb, L.R. Hope, A.E. Nicholson, and K. Axnick. 2004. Varieties of causal intervention. In *PRICAI*, pp. 322–331.

J. Pearl. 1995. Causal diagrams for empirical research. *Biometrika*, 82:669–688.

J. Peters, J.M. Mooij, D. Janzing, and B. Schölkopf. 2011. Identifiability of causal graphs using functional models. In *UAI*, pp. 589–598.

D.J. Rose, R.E. Tarjan, and G.S. Lueker. 1976. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5(2):266–283.

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

P. Spirtes, C.N. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search.*

S. Tong and D. Koller. 2001. Active learning for structure in Bayesian networks. In *IJCAI*, vol 17, pp. 863–869.

T. Verma and J. Pearl. 1990. On the equivalence of causal models. In *UAI*, pp. 220–227.