On Finding a Small Set of Atomic Interventions to Identify a Causal Model

Dylan Fridman

June 2025

Abstract

In this expository paper, we present one of the main results of *Verification and Search Algorithms for Causal DAGs*: that searching for the correct causal graph given its essential graph can be done using no more than $O(\log n)$ times the minimum number of atomic interventions needed to verify that it is the ground truth [1].

1 Introduction

Even if we assume faithfulness[1] and no unobserved variables, it is well-known that a causal model might not be able to be identified from observational data alone: some causal models require interventions for us to be able to identify them. It turns out that figuring out a small set of interventions to identify the true causal model is not a trivial task.

In *Verification and Search Algorithms for Causal DAGs* [1], Choo et al. show that the problem of finding the correct causal model given an essential graph is not significantly harder than it is to *verify* that a given DAG is the correct causal model. They give an algorithm for which the "search problem" requires only $O(\log n)$ times more atomic interventions than the minimum number needed to verify the model.

In terms of this verification number, their algorithm is asymptotically tight. Let $L_n^{(k)}$ be the DAG on n nodes $\{v_i\}_{i\in[n]}$ for which $v_i \leftarrow v_{i+1}$ for $i \in \{1, 2, ..., k-1\}$ and $v_i \rightarrow v_{i+1}$ for $k \in \{k, k+1, ..., n\}$. Note that the essential graph $\mathcal{E}(L_n^{(k)})$ has no directed edges. Now, if we know that $L_n^{(k)}$ is the true causal model, verifying it is very easy: a single atomic intervention on the root will suffice. However, if we are only given the essential graph, we will need at least $O(\log n)$ atomic interventions, since the problem is equivalent to a binary search.

Figure 1: For line graphs, verifying the causal model requires only one atomic intervention, but searching for it requires $O(\log n)$ interventions.

[1]Faithfulness refers to the idea that if two variables are independent in the distribution, then they are d-separated in the causal model.

Thus, the problem of searching for can be harder than verifying it by a factor of $O(\log n)$. In this paper, we will present the main ideas behind their algorithm and prove that it indeed requires at most $O(\log n \cdot \nu_1(G))$ atomic interventions, where $\nu_1(G)$ is the minimum number of atomic interventions needed to verify that a given DAG G is the true causal model.

2 Preliminaries

The main idea behind their algorithm is the following. We know that intervening on the endpoint of an undirected edge in the essential graph will orient it [2]. Now, start by ignoring the directed edges and for each component of the resulting graph, find a small $\frac{1}{2}$-separator. Repeat this process until all edges are oriented.

For the purpose of this expository paper, we are only going to work with atomic interventions. That is, interventions that consist of a single node. For a DAG $G = (V, E)$ and a set of interventions $\mathcal{I} \subset V$, we let $\mathcal{E}(G)$ be its essential graph and $\mathcal{E}_{\mathcal{I}}(G)$ be its $\mathcal{I}$-essential graph, i.e., the graph representing the equivalence class of all graphs that are Markov equivalent regardless of which intervention (or none) we apply on the graph.[2]

Given a graph G with both directed and undirected edges, we call its *chain components* $\mathrm{CC}(G)$ the connected components that we obtain after removing all directed edges. We know that the chain components of $\mathcal{E}_{\mathcal{I}}(G)$ for any $\mathcal{I}$ (including $\mathcal{I} = \emptyset$) are chordal, i.e., every cycle of length ≥ 4 has a an edge that is not part of the cycle but connects two vertices of the cycle [2]. Another key concept is that of an α-clique separator:

Definition 2.1 (α-separator and α-clique separator)**.** Let A, B, C be a partition of the vertices V of a graph $G = (V, E)$. We say that C is an α-separator if no edge joins a vertex in A with a vertex in B and $|A|, |B| \leq \alpha \cdot |V|$. We say C is an α-clique separator if it is an α-separator and a clique.

It turns out that, for chordal graphs, we can obtain $\frac{1}{2}$-clique separators in polynomial time:

Theorem 2.1. *Let $G = (V, E)$ be a chordal graph with $|V| \geq 2$ and p vertices in its largest clique. Then there exists a $\frac{1}{2}$-clique separator C with $|C| \leq p - 1$.*

3 Result

We can now state the result a more formally:

Theorem 3.1. *Let G be a DAG with n vertices. Then, Algorithm 1, given its essential graph $\mathcal{E}(G)$, finds G using at most $O(\log n \cdot \nu_1(G))$ atomic interventions.*

Algorithm 1 is relatively simple: it repeatedly finds $\frac{1}{2}$-clique separators of the chain components of the $\mathcal{I}$-essential graph (where $\mathcal{I}$ is the set of interventions so far) using Theorem 2.1 and the fact that they are chordal. Then, it intervenes in the vertices of the separator. This means that, after each step, the size of the largest chain component graph is halved.

[2]We can think of applying an intervention as disconnecting the intervened node from its parents.

Algorithm 1 Search algorithm via graph separators

Input: Essential graph $E(G^*)$, **Output:** Fully oriented graph $G \in [G^*]$
Initialize $i \leftarrow 0$, $\mathcal{I}_0 \leftarrow \emptyset$
while $\mathcal{E}_{\mathcal{I}_i}(G^*)$ has undirected edges **do**
    For each $H \in \mathrm{CC}(\mathcal{E}_{\mathcal{I}_i}(G^*))$ with $|H| \geq 2$, find a 1/2-clique separator K_H using Theorem 2.1
    Define $C_i \leftarrow \bigcup_{H \in \mathrm{CC}(\mathcal{E}_{\mathcal{I}_i}(G^*)), |H| \geq 2} K_H$ as the atomic intervention set
    $i \leftarrow i + 1$, intervene on C_i to obtain $\mathcal{E}_{\mathcal{I}_i}(G^*)$, and update $\mathcal{I}_i \leftarrow \mathcal{I}_{i-1} \cup C_i$

When the size of the largest chain component is 1, we know that all edges have been oriented. Thus, the algorithm terminates in $O(\log n)$ iterations. Furthermore, each iteration requires at most $O(\nu_1(G))$ atomic interventions, the total number of interventions is $O(\log n \cdot \nu_1(G))$. This follows from the following lemma:

Lemma 1. *Let $G^* = (V, E)$ be a DAG, G another DAG Markov equivalent to G^*, and $\mathcal{I} \subset V$ a set of atomic interventions. Then,*

$$\nu_1(G) \geq \sum_{H \in \mathrm{CC}(\mathcal{E}_{\mathcal{I}}(G^*))} \left\lfloor \frac{\omega(H)}{2} \right\rfloor$$

where $\omega(H)$ is the size of the largest clique in H.

Hence, this shows that the number of atomic interventions used by Algorithm 1 is at most $O(\log n \cdot \nu_1(G))$.

4 Conclusion

In this paper, we presented the main ideas behind the algorithm proposed in *Verification and Search Algorithms for Causal DAGs* [1]. The algorithm is a search algorithm that uses $\frac{1}{2}$-clique separators to orient the edges of the essential graph, and it requires at most $O(\log n \cdot \nu_1(G))$ atomic interventions to find the true causal model given its essential graph.

This result is significant because it shows that the search problem is not significantly harder than the verification problem, at least in terms of the number of atomic interventions needed. Furthemore, it closes the gap by matching the lower bound given by the example of line graphs, where the search problem requires $O(\log n)$ times more interventions than the verification problem.

References

[1] Choo, D., Shiragur, K., Bhattacharyya, A. (2022). *Verification and Search Algorithms for Causal DAGs.* arXiv preprint arXiv:2206.15374.

[2] Alain Hauser and Peter Bühlmann. (2012). *Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs.* arXiv preprint arXiv:1104.2808