SNAP: Sequential Non-Ancestor Pruning for Targeted Causal Effect Estimation With an Unknown Graph

Mátyás Schubert[1] **Tom Claassen**[2] **Sara Magliacane**[1]

[1]Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
[2]Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

Abstract

Causal discovery often serves as a precursor to causal effect estimation, but it can be computationally demanding due to the number of conditional independence tests involved. If we are interested in estimating only the causal effects on a small subset of the measured variables, many of these tests may be unnecessary. Existing methods addressing this issue often have strong assumptions about the causal relations between variables. In this paper, we consider *targeted causal effect estimation with an unknown graph*, a task that focuses on identifying the causal effect between multiple target variables. This task combines causal discovery and effect estimation, aligning the discovery objective with the effects to be estimated. We show that the non-ancestors of the target variables are unnecessary to estimate the causal effects between the targets. We sequentially identify and prune these non-ancestors during the process of existing algorithms. Our results show that our approach substantially reduces the number of tests without compromising the quality of causal effect estimations.

1 MOTIVATION

Causal inference [Pearl, 2009] is fundamental to our scientific understanding and practical decision-making. In many settings, we do not know the causal relations between the variables, which we can learn with *causal discovery* methods [Glymour et al., 2019]. These methods can be computationally demanding for large numbers of variables. In many cases, we are only interested in estimating the causal effects between a small subset of variables, which would not require recovering the causal graph over all variables.

We formalize this setting as *targeted causal effect estimation with an unknown graph*, a task that focuses on identifying the causal effects $P(T_i|do(T_j))$ between pairs of target variables $T_i, T_j \in \mathbf{T}$, where $\mathbf{T}$ is a small subset of all variables $\mathbf{V}$ in a *computationally efficient way*. In this setting, we assume that we do not have access to the true causal graph, but to the joint observational distribution p over $\mathbf{V}$ that is Markov and faithful to the true causal graph, and causally sufficient, i.e., no unobserved confounders or selection bias.

Under these assumptions, we can use constraint-based causal discovery algorithms [Spirtes et al., 2000] to identify the Markov equivalence class (MEC) of the causal graph Verma and Pearl [1990], represented by a mixed graph, called the complete partially directed acyclic graph (CPDAG). The CPDAG can then be used to identify valid adjustment sets for causal effect estimation [Perković et al., 2015]. However, discovering the CPDAG over all variables can scale poorly in terms of conditional independence (CI) tests for large numbers of nodes [Mokhtarian et al., 2021].

Local causal discovery methods [Wang et al., 2014, Gupta et al., 2023] aim to address this issue by identifying the parent adjustment set of a single treatment-outcome pair. Thus, these are not designed to handle more than two targets with unknown causal relations. Watson and Silva [2022] developed an algorithm to discover the causal relations between multiple targets, which they call *foreground variables*, but assume that the other variables, the *background variables*, are all non-descendants of the target variables.

To fill this gap, we propose Sequential Non-Ancestor Pruning (SNAP), a principle to efficiently identify the causal relations and valid adjustment sets between targets, without requiring any assumptions about the causal graph.

2 METHOD

SNAP improves the computationally efficiency of causal discovery by avoiding CI tests that do not contribute to estimating the causal effects on the target variables $\mathbf{T}$. To do this, we show that the non-ancestors of the targets $N(\mathbf{T})$ do not help in orienting paths between the targets and in their

potential adjustment sets. Furthermore, they are not part of statistically efficient adjustment sets, such as the parent, canonical [Perković et al., 2015] or optimal [Henckel et al., 2022] adjustment sets. Thus, performing causal discovery only on $\mathbf{V} \setminus N(\mathbf{T})$ also enables the estimation, while potentially requiring much fewer CI tests. While $N(\mathbf{T})$ is not available before causal discovery, we can often obtain partial information about it *during* the process. Driven by this insight, we aim to progressively identify and remove $N(\mathbf{T})$ throughout the process of a causal discovery algorithm.

We demonstrate this approach by modifying the PC algorithm [Spirtes et al., 2000]. The original PC algorithm starts with the skeleton search, which identifies the adjacencies of the CPDAG by iteratively performing sets of CI tests with increasing order $i = 0, 1, ...$, describing the size of the conditioning set. This is followed by the orientation v-structures and the exhaustive application of Meek rules [Meek, 1995].

We propose the Sequential Non-Ancestor Pruning (SNAP) algorithm in which we modify the skeleton search of PC by orienting v-structures and applying Meek rules after each set of CI tests at a given order i, identifying a subset of $N(\mathbf{T})$ with these orientations, and continuing the search only over the remaining variables. We show that SNAP is sound at every step of the skeleton search, i.e., it never removes any ancestors of the targets. Thus, SNAP can be run up to any order k to obtain a subset of $N(\mathbf{T})$. SNAP is described in Algorithm 1, where the extra steps are colored in blue, while the other steps resemble the classical PC algorithm. Since SNAP considers fewer and fewer variables with every order, it uses fewer higher order CI tests.

Algorithm 1 Sequential Non-Ancestor Pruning - SNAP(k)

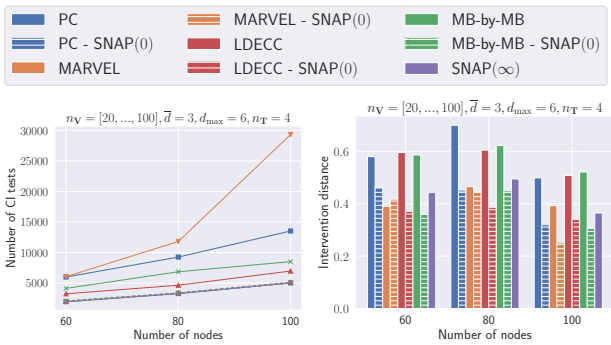
Require: Data $\mathcal{D}$, Targets $\mathbf{T} \subseteq \mathbf{V}$, Maximum order k
$\hat{G} = (\hat{\mathbf{V}}, \hat{\mathbf{E}}) \leftarrow$ fully connected undirected graph
for $i \in 0..k$ **do**
  for $(X, Y) \in \hat{\mathbf{E}}$ **do**
    for $\mathbf{S} \subseteq Adj_{\hat{G}^i}(X) \setminus \{Y\}$ s.t. $|\mathbf{S}| = i$ **do**
      if $X \perp\!\!\!\perp Y | \mathbf{S}$ in $\mathcal{D}$ **then**
        Delete the edge $X - Y$ from $\hat{\mathbf{E}}$
        break
  $\hat{G}^i \leftarrow$ Orient v-structures and Meek rules on $\hat{G}$
  $\hat{G}^i \leftarrow$ Remove edges from conflicting v-structures
  Remove all $V \in \hat{\mathbf{V}}$ that do not have a possibly directed path to any $T \in \mathbf{T}$ in $\hat{G}^i$
return $\hat{V}$, $\hat{G}^k$

SNAP can be used as a pre-filtering method for standard causal discovery algorithms, which are then applied only on the remaining variables $\hat{V}$ that were not identified to be in $N(\mathbf{T})$. We denote this filtering process as SNAP(k), where k is the maximum order of conditional independence tests used by SNAP. Additionally, we can also run SNAP until completion, i.e. $k = |\mathbf{V}|$, by adding an extra step at the end that orients the v-structures and uses the Meek rules on $\hat{G}^k$ restricted to $\hat{V}$. We denote this method as SNAP(∞).

3 RESULTS

We evaluate SNAP(∞), PC [Spirtes et al., 2000], MARVEL [Mokhtarian et al., 2021], MB-by-MB [Wang et al., 2014] and LDECC [Gupta et al., 2023], and their combination with SNAP(0). As local algorithms take as input a single target, we apply them on all targets separately and aggregate the results. Since they require oracle knowledge of the causal relations between the targets, we provide this as input.

In Figure 1a, we report results for linear Gaussian data. We sample 1000 data points according to 100 random causal graphs with edge coefficients in $[-3, -0.5] \cup [0.5, 3]$ and unit variance, expected degree of $\overline{d} = 3$, maximum degree of $d_{\max} = 6$ and different numbers of nodes $n_{\mathbf{V}}$. We sample $n_{\mathbf{T}} = 4$ *identifiable* targets, such that the causal effects are identifiable pairwise between all pairs of targets from the corresponding true CPDAG, and each target is an ancestor or descendant of at least one other target.



(a) Number of CI tests (b) Intervention distance

Figure 1: Number of CI tests performed (a) and intervention distance (b) over graphs with different number of nodes.

Fig. 1a shows the number of partial correlation tests with $\alpha = 0.05$ over graphs with different numbers of nodes. Our results show that the prefiltering with SNAP(0) (represented by the dashed lines) reduces the number of tests for each of the causal discovery methods, as can be seen by the overlapping lines at the bottom of the graph. Fig. 1b shows that in this setting prefiltering with SNAP(0) with canonical adjustment also reduces the intervention distance, i.e. the distance between the predicted and true $P(T_i|do(T_j))$. In these experiments, to make the methods comparable across all of the graphs, we consider $P(T_i|T_j)$ as the causal effect, when a method outputs that the effect is not identifiable.

In additional preliminary experiments, we find that vanilla MARVEL performs on par with SNAP methods in terms of CI tests when using oracle CI tests and binary tests. However, MARVEL fails on binary data from larger graphs due to it using total conditioning [Pellet and Elisseeff, 2008].

References

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10, 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00524.

Shantanu Gupta, David Childers, and Zachary Chase Lipton. Local causal discovery for estimating causal effects. In *Conference on Causal Learning and Reasoning*, pages 408–447. PMLR, 2023.

Leonard Henckel, Emilija Perković, and Marloes H Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):579–599, 2022.

Christopher Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 403–410, 1995.

Ehsan Mokhtarian, Sina Akbari, AmirEmad Ghassami, and Negar Kiyavash. A recursive markov boundary-based approach to causal structure learning. In *The KDD'21 Workshop on Causal Discovery*, pages 26–54. PMLR, 2021.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jean-Philippe Pellet and André Elisseeff. Using markov blankets for causal structure learning. *Journal of Machine Learning Research*, 9(7), 2008.

Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H Maathuis. A complete generalized adjustment criterion. In *Uncertainty in Artificial Intelligence-Proceedings of the Thirty-First Conference (2015)*, pages 682–691. AUAI Press, 2015.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '90, page 255–270, USA, 1990. Elsevier Science Inc. ISBN 0444892648.

Changzhang Wang, You Zhou, Qiang Zhao, and Zhi Geng. Discovering and orienting the edges connected to a target variable in a dag via a sequential local learning approach. *Computational statistics & data analysis*, 77:252–266, 2014.

David S Watson and Ricardo Silva. Causal discovery under a confounder blanket. In *Uncertainty in Artificial Intelligence*, pages 2096–2106. PMLR, 2022.