A structure independent algorithm for causal discovery

Tom Claassen and Tom Heskes *

Radboud University Nijmegen - Intelligent Systems
Heyendaalseweg 135, 6525 AJ Nijmegen - The Netherlands

Abstract. We present two inference rules, based on so called *minimal conditional independencies*, that are sufficient to find *all* invariant arrowheads in a single causal DAG, even when selection bias may be present. It turns out that the set of seven graphical orientation rules that are usually employed to identify these arrowheads are, in fact, just different instances/manifestations of these two rules. The resulting algorithm to obtain the definite causal information is elegant and fast, once the (often surprisingly small) set of minimal independencies is found.

1   Introduction

Causal discovery remains at the heart of most scientific research to date. Understanding which variables in a causal system influence which other is crucial for predicting the effects of actions and policies. Sometimes, it is very important to know that a certain variable is *not* the cause of another. Correctly identifying such relations from data is the focus of this article.

A popular and intuitive way of representing a causal system is in the form of a directed acyclic graph (DAG). A causal DAG $\mathcal{G}_C$ is a graphical model where the arrows represent direct causal interactions between variables in a system [1, 2]. There is a causal relation $X \Rightarrow Y$, iff there is a directed path from X to Y in $\mathcal{G}_C$. The causal Markov condition links the structure of a causal graph to observed conditional independencies $X \perp\!\!\!\perp Y \mid \mathbf{Z}$. For details on probabilistic graphical model concepts and terminology, the reader is referred to [1, 3].

When some variables in the causal DAG are hidden, or when there is possible selection bias [4], the independence relations between the observed variables can be represented in the form of a maximal ancestral graph (MAG) [5]. The (complete) partial ancestral graph (cPAG) represents all invariant features that characterize the equivalence class $[\mathcal{G}]$ of such a MAG, with a tail '−' or arrowhead '>' mark on an edge, iff it is invariant in $[\mathcal{G}]$, otherwise it has a circle mark '∘', see [6]. Tails in a PAG are associated with identifiable direct causal relations, and arrowheads with the absence thereof, [2, 7, 8]. Fig.1 illustrates the relation between these three types of graphs.

Recently, Claassen and Heskes [7] showed how *minimal* conditional independencies, $[X \perp\!\!\!\perp Y \mid \mathbf{Z}]$, indicating that no proper subset $\mathbf{Z}' \subsetneq \mathbf{Z}$ can make X and Y independent, could be employed to infer causal relations from multiple models. The method was sound, but not complete. After section 2 offers a glimpse of
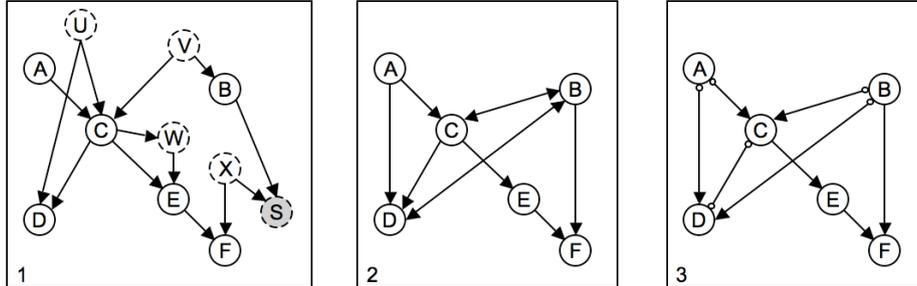
Fig. 1: 1) Causal DAG $\mathcal{G}_C$ over 11 nodes (dashed = hidden, grayed S = selection variable); 2) corresponding causal MAG over observed nodes; 3) cPAG.

the state of the art approach to constraint-based causal discovery (in particular, how to obtain all invariant arrowheads), section 3 makes a first step towards completeness, by showing that the two rules behind this method are also sufficient to cover all invariant arrowheads in a single model. Section 4 puts them to work in an algorithm, and section 5 discusses a number of extensions.

2   Learning causal models from data

The challenge of causal discovery is how to identify all these invariant features from a given data set, to determine which variables do or do not have a directed path to which others in the underlying causal DAG.

The famous Fast Causal Inference (FCI) algorithm [3] was one of the first algorithms that was able to validly infer causal relations from conditional independence statements in the large sample limit, even in the presence of latent and selection variables. It consists of an efficient search for a conditional independence between each pair of variables to identify the skeleton of the underlying causal MAG, followed by an orientation stage to identify invariant tail and arrowhead marks. It was shown to be sound [4], although not yet complete. Ali et al. [9] showed that a set of seven graphical orientation rules was sufficient to identify all invariant arrowheads in the equivalence class $[\mathcal{G}]$, given a single MAG $\mathcal{G}$. Algorithm 1, below, shows the implementation of these rules in the context of the FCI algorithm in [6], where we ignore the details behind the initial adjacency search. (When starting from data instead of a MAG, each rule is formulated as an equivalent set of (in)dependence statements.)

3   Invariant arrowheads and minimal independencies

This section derives the main result of this paper: that the seven graphical orientation rules in algorithm 1 are, in fact, different manifestations of just two(!) rules. To prove this, we first need to generalize the result in theorem 1 of [7] to

Input : independence oracle, fully $\circ\!-\!\circ$ connected graph $\mathcal{P}$ over $\mathbf{V}$
Output : PAG $\mathcal{P}$
1: **for all** $\{X, Y\} \in \mathbf{V}$ **do**
2: search *in some clever way* for a $X \perp\!\!\!\perp Y \mid \mathbf{Z}$, if found
3:   $\mathcal{R}0a$: eliminate $X \not\!\times Y$ from $\mathcal{P}$
4:   record $Sepset(X, Y) \leftarrow \mathbf{Z}$
5: **end for**
6: $\mathcal{R}0b$: orient $X \ast\!\!\rightarrow Z \leftarrow\!\!\ast Y$, iff $X \not\!\times Y$ and $Z \notin Sepset(X, Y)$
7: **repeat**
8: $\mathcal{R}1$: orient $Z \longrightarrow Y$, if $X \not\!\times Y$ and $X \ast\!\!\rightarrow Z$
9: $\mathcal{R}2a$: orient $Z \ast\!\!\rightarrow Y$, if $Z \longrightarrow X \ast\!\!\rightarrow Y$
10: $\mathcal{R}2b$: orient $Z \ast\!\!\rightarrow Y$, if $Z \ast\!\!\rightarrow X \longrightarrow Y$
11: $\mathcal{R}3$: orient $W \ast\!\!\rightarrow Z$, if $X \ast\!\!\rightarrow Z \leftarrow\!\!\ast Y$, $X \ast\!\!-\!\circ W \circ\!-\!\ast Y$ and $X \not\!\times Y$
12: if $\langle X, Z_1, \ldots, Z_k, Z, Y \rangle$ is a discriminating path for Z, then
13:   $\mathcal{R}4a$: orient $Z \longrightarrow Y$, iff $Z \in Sepset(X, Y)$
14:   $\mathcal{R}4b$: orient $Z_k \longleftrightarrow Z \longleftrightarrow Y$, iff $Z \notin Sepset(X, Y)$
15: **until** no new orientations found

Algorithm 1: FCI for invariant arrowheads

allow for the possibility of selection bias.

Lemma 1. Let X, Y, $\mathbf{Z}$ and W be four disjoint (sets of) observed nodes in a causal DAG $\mathcal{G}_C$, and $\mathbf{S}$ be a set of (hidden) selection nodes, then:

- a *minimal* conditional independence $[X \perp\!\!\!\perp Y \mid \mathbf{Z}]$ implies directed paths $Z \Rightarrow \{X/Y/S\}$ from every $Z \in \mathbf{Z}$ to X and/or Y and/or $S \in \mathbf{S}$ in $\mathcal{G}_C$,[1]
- a conditional dependence $X \not\!\perp\!\!\!\perp Y \mid \{\mathbf{Z} \cup W\}$, created by W from $[X \perp\!\!\!\perp Y \mid \mathbf{Z}]$, implies that there are no directed paths from W to X, Y, $\mathbf{Z}$ or $\mathbf{S}$ in $\mathcal{G}_C$.

Proof. Analogous to theorem 1 in [7], but now accounting for the fact that selection can induce additional dependencies. For details see [8]. $\square$

Together, the two rules allow to infer causal relations, even in the presence of selection bias: find a minimal conditional independence $[X \perp\!\!\!\perp Y \mid \mathbf{Z}]$, and eliminate $Z \not\!\Rightarrow X$ and $Z \not\!\Rightarrow \mathbf{S}$ by a conditional dependence $X \not\!\perp\!\!\!\perp U \mid \mathbf{W} \cup Z$ created by some $Z \in \mathbf{Z}$, to infer $Z \Rightarrow Y$. We can now state the main theorem as

Theorem 2. In a PAG $\mathcal{G}$, all invariant arrowheads $X \ast\!\!\rightarrow Y$ are instances of

rule (1): $U \not\!\perp\!\!\!\perp V \mid \mathbf{W} \cup Y$, created by Y from a *minimal* $[U \perp\!\!\!\perp V \mid \mathbf{W}]$, with $X \in \{U, V, \mathbf{W}\}$, or

rule (2): a minimal $[Y \perp\!\!\!\perp Z \mid \mathbf{W} \cup X]$, with an arrowhead at $Z \ast\!\!\rightarrow X$ from either rule (1) or rule (2).

---

[1] Many thanks to Peter Spirtes for pointing out that a similar observation was already made in [4] (corollary to lemma 14), although it was only used to prove correctness of the FCI-algorithm and never used as an orientation rule.

Proof sketch. Both rules are sound, as they are direct applications of Lemma 1. The proof that they are also complete follows by induction on the graphical orientation rules $\mathcal{R}0b - \mathcal{R}4b$, showing that none of them introduces a violation of Theorem 2. As these rules are sufficient for arrowhead completeness, it follows that the theorem holds for all invariant arrowheads. For the full proof, see [8]. $\square$

4 An algorithm for arrowhead completeness

We can use the rules in Theorem 2 in an algorithm that uncovers all invariant arrowheads directly from the minimal conditional independencies (and subsequent dependencies) found, without having to refer to the structure of the graph. In fact, we do not even need to find all minimal independencies, see [8]:

Lemma 3. Finding a *single* minimal independence for each pair of nodes $\{X, Y\}$ in the graph (if it exists) is sufficient to orient all invariant arrowheads.

Fortunately, the standard implementation of the FCI-algorithm already finds only minimal sets, as it looks for separating sets $\mathbf{Z}$ of increasing size. Finally, rule (2) is guaranteed to find only definitely causal tails, even with selection bias:

Lemma 4. The transitive closure of the invariant arcs found by rule (2) all correspond to identifiable, definite causal relations in the underlying causal DAG.

> **Input** : independence oracle, fully $\circ\!\!-\!\!\circ$ connected graph $\mathcal{P}$ over $\mathbf{V}$
> **Output** : PAG $\mathcal{P}$, causal relations matrix $\mathbf{M}_C$
> 1: **for all** $\{X, Y\} \in \mathbf{V}$ **do**
> 2:   search *in some clever way* for a minimal $[X \perp\!\!\!\perp Y \mid \mathbf{Z}]$, if found
> 3: $\mathbf{S}_{CI} \leftarrow$ triple $(X, Y; Z), \forall Z \in \mathbf{Z}$
> 4: $\mathbf{R}_{CD} \leftarrow (\{X/Y/\mathbf{Z}\}; W), \forall W : X \not\!\perp\!\!\!\perp Y \mid \mathbf{Z} \cup W$
> 5: **end for**
> 6: $\mathbf{M}_C \leftarrow (Z \not\Rrightarrow Y)$, iff $(Y; Z) \in \mathbf{R}_{CD}$ $\triangleright$ rule (1)
> 7: **repeat**
> 8: $\mathbf{M}_C \leftarrow (Z \Rightarrow Y)$, if $(X, Y; Z) \in \mathbf{S}_{CI}$ and $Z \not\Rrightarrow X \in \mathbf{M}_C$ $\triangleright$ rule (2)
> 9: **until** no new information found
> 10: $\mathbf{M}_C \leftarrow$ transitive closure of all $(Z \Rightarrow Y) \in \mathbf{M}_C$
> 11: $\mathcal{P} \leftarrow$ eliminate $X \not\!\!\Rightarrow\!\!\Leftarrow Y$, iff $(X, Y; *) \in \mathbf{S}_{CI}$
> 12: $\mathcal{P} \leftarrow$ orient $X \!-\!* Y$, iff $(X \Rightarrow Y) \in \mathbf{M}_C$
> 13: $\mathcal{P} \leftarrow$ orient $X *\!\!\rightarrow Y$, iff $(Y \not\Rrightarrow X) \in \mathbf{M}_C$

Algorithm 2: Algorithm for invariant arrowheads

Algorithm 2 provides an implementation. The first part, lines $1-5$, is essentially the same as before, except that when a (minimal) independency is found, it also records which nodes destroy this dependency in $\mathbf{R}_{CD}$ (line 4). The actual identification part, lines $6 - 10$, transfers the recorded information in the $\mathbf{R}_{CD}$ and

$\mathbf{S}_{CI}$ structures directly to the causal matrix $\mathbf{M}_C$, where one instance of rule (2) may trigger another (causal chain). The final part, lines $11-13$, simply map the recorded explicit causal information to tails, arrowheads and edge eliminations in the equivalent PAG representation.

5   Experimental results and discussion

We verified the algorithm and tested its behaviour on a large set of randomly generated causal DAGs with different edge densities and random numbers of hidden and/or selection nodes. For each graph, the 'true' causal DAG was used to function as the independence oracle. We computed all invariant arrowheads, and compared with the cPAG calculated over the observed nodes. We also recorded the number of (minimal) conditional (in)dependencies.

In all cases, all invariant arrowheads were correctly identifed, as predicted by Theorem 2, and also a significant number of invariant tails. The computational complexity of the orientation part in algorithm 2 is low compared to algorithm 1, as we do not have to check for discriminating paths, see also [10]. However, the overall complexity in both algorithms is dominated by the initial search for (minimal) conditional independencies, not the number found. Tabel 1 gives an indication of the number of these independencies as a function of size for graphs with average edge density.

nodes	6	8	10	12	14
cond.ind.	35.2	369	2820	18700	47300
min. cond.ind.	6.2	18.0	37.6	67.6	107
cond.dep.	5.0	18.0	45.1	92.5	165

Table 1: Average nr. of (minimal) cond.(in)dep. as a function of graph size

From a theoretical viewpoint, the structure independent characterization of invariant arrowheads is quite interesting, and raises the question if a similar approach is viable for invariant tails as well. Preliminary results suggest this is indeed the case, but (probably) requires two more rules, similar to Theorem 2.

Perhaps the most promising aspect of the algorithm is its potential to address the lack of robustness and flexibility that is often associated with constraint-based methods for real world data sets, as compared to, for instance, Bayesian scoring methods for causal discovery [11, 12]. As graph based orientation rules rely on categorical independence decisions, one erroneous CI-test may lead to a sequence of false orientations, even inconsistencies, without any trace of this ambiguity in the final output graph. Our result shows that constraint-based methods are not necessarily susceptible to this kind of instability: the structure independent conditions in Theorem 2 can provide a direct measure of the degree of certainty of a particular causal conclusion, similar to Bayesian methods. It is also possible to use multiple estimates from different node combinations to obtain a more robust valuation and detect possible inconsistencies. However, a full treatment goes far beyond the scope of this article.

Finally, from the derivation of the algorithm it is clear that, even if not all independencies are found, then still all invariant tails and arrowheads identifed by algorithm 2 remain valid. Therefore, we can bring the orientation part, lines $6-10$, *inside* the CI-loop, updating each time a minimal conditional independency is found, to effectively turn it into an anytime algorithm.

6 Conclusion

We have shown that all invariant arrowheads in the equivalence class of a causal DAG with latent and selection variables (cPAG) are instances of just two rules, that both start from an observed minimal conditional independence. These arrowheads, $X *\!\!\to Y$, represent all detectable information of the form 'Y does not cause X' in a data set in the large sample limit.

We applied the rules in a straightforward and efficient algorithm, that is capable of extracting identifiable causal relations from independence relations, when selection bias may be present. The fact that the algorithm does not rely on the graphical structure of the PAG opens up a number of interesting possibilities for further extensions, including more robust/direct, Bayesian estimates for the likelihood of individual causal relations, and detection of inconsistencies.

We are currently working on two additional rules to cover invariant tails as well. Ultimate goal is to come to some form of completeness result in the multiple model domain for the approach taken in [7].

References

[1] J. Pearl. *Causality: models, reasoning and inference*. Cambridge University Press, 2000.

[2] J. Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437 − 1474, 2008.

[3] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. The MIT Press, Cambridge, Massachusetts, 2nd edition, 2000.

[4] P. Spirtes, C. Meek, and T. Richardson. An algorithm for causal inference in the presence of latent variables and selection bias. In *Computation, Causation, and Discovery*. 1999.

[5] T. Richardson and P. Spirtes. Ancestral graph Markov models. *Ann. Stat.*, 30(4), 2002.

[6] J. Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17), 2008.

[7] T. Claassen and T. Heskes. Causal discovery in multiple models from different experiments. In *Advances in Neural Information Processing Systems 23*, pages 415–423. 2010.

[8] T. Claassen and T. Heskes. Arrowhead completeness from minimal conditional independencies. Technical report, Faculty of Science, Radboud University Nijmegen, 2010.

[9] R.A. Ali, T. Richardson, P. Spirtes, and J. Zhang. Towards characterizing markov equivalence classes for directed acyclic graphs with latent variables. In *Proceedings of the 21th Conference on Uncertainty in Artificial Intelligence*, pages 10–17. AUAI Press, 2005.

[10] R Ali, T. Richardson, and P. Spirtes. Markov equivalence for ancestral graphs. *The Annals of Statistics*, 37(5B):2808–2837, 2009.

[11] D. Heckerman, C. Meek, and G. Cooper. A Bayesian approach to causal discovery. In *Computation, Causation, and Discovery*, pages 141–166. 1999.

[12] D. Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(3):507–554, 2002.