A Weaker Faithfulness Assumption based on Triple Interactions (Supplementary Material)

Alexander Marx[1,2] **Arthur Gretton**[3] **Joris M. Mooij**[4]

[1]CISPA Helmholtz Center for Information Security, Saarland University, Saarbrücken, Germany
[2]Max Planck Institute for Informatics, Saarland University, Saarbrücken, Germany
[3]Gatsby Unit, University College London, London, United Kingdom
[4]Korteweg-de Vries Institute, University of Amsterdam, Amsterdam, The Netherlands

SUPPLEMENTARY MATERIAL

S.1 EXAMPLE 2 IN DETAIL

As described in Section 6, we can generate a DAG of the form $X \to Y \leftarrow Z$ and $W \to Y$ s.t. X, Y and Z form a minimal unfaithful triple and $W \not\perp\!\!\!\perp_P Y$ as follows. We generate X, Z, W and E independently, with X and Z as fair coins, W as a coin with $P(W = 1) = p$, where $0 < p < 1$ and E (the noise variable) as a biased coin with $P(E = 1) = q$, $0 < q < \frac{1}{2}$. With $q > 0$, we ensure that the function is non-deterministic. Further, we generate Y as

$$Y := ((X \oplus Z) \wedge W) \oplus E \ .$$

We will obtain that $P(Y = 1) = q + \frac{p}{2} - pq$. Further, we can calculate that $P(X = 1, Y = 1) = \frac{1}{2}P(Y = 1) = P(X = 1) \cdot P(Y = 1)$. Also, $P(X = 1, Y = 0) = P(X = 1) \cdot P(Y = 0)$, which means that they are marginally independent. The same holds for Z and Y. If we calculate the probability for all three variables, we get that $P(X = 0, Z = 1, Y = 1) = \frac{p+q-2pq}{4}$ and $P(X = 0, Z = 1) \cdot P(Y = 1) = \frac{1}{4}P(Y = 1)$. Hence, we need to solve

$$\begin{aligned} P(X = 0, Z = 1, Y = 1) &= P(X = 0, Z = 1) \cdot P(Y = 1) \\ \Leftrightarrow p + q - 2pq &= q + \frac{p}{2} - pq \\ \Leftrightarrow p - pq &= \frac{p}{2} \ . \end{aligned}$$

The only solutions are $p = 0$ or $q = \frac{1}{2}$, which we excluded. Hence, $Y \not\perp\!\!\!\perp_P \{X, Z\}$ and by weak union also $Y \not\perp\!\!\!\perp_P X \mid Z$, as well as $Y \not\perp\!\!\!\perp_P Z \mid X$. Since we know by assumption that $X \perp\!\!\!\perp_P Z$ we can conclude from Lemma 1 that also $X \not\perp\!\!\!\perp_P Z \mid Y$, which means that $\{X, Y, Z\}$ from a minimal unfaithful triple since W will also not cancel out any of these conditional dependencies. Next, we also find that $W \not\perp\!\!\!\perp_P Y$, since $P(W = 1, Y = 1) = \frac{p}{2}$, which is only equal to $P(W = 1) \cdot P(Y = 1)$, if $p = 0$, $p = 1$ or $q = \frac{1}{2}$, which we excluded, and hence $W \not\perp\!\!\!\perp_P Y$. Last, we need to show that $X \not\perp\!\!\!\perp_P W \mid \{Y, Z\}$ and that $Z \not\perp\!\!\!\perp_P W \mid \{X, Y\}$. We can write

$$P(X, W \mid Y, Z) = \frac{P(X, W, Y, Z)}{P(Y, Z)} \ .$$

To show conditional dependence, this value has to be different from $P(X \mid Y, Z) \cdot P(W \mid Y, Z)$. Consider the case where all variables are equal to one. Hence, we get that

$$\begin{aligned} P(X = 1, W = 1, Y = 1, Z = 1) &= \frac{pq}{4} \ , \\ P(X = 1, Y = 1, Z = 1) &= \frac{q}{4} \ , \\ P(W = 1, Y = 1, Z = 1) &= \frac{p}{4} \ . \end{aligned}$$

Since we know that $P(Y = 1, Z = 1) = P(Y = 1)/2$, we thus need to solve

$$pq = \frac{pq}{2P(Y = 1)} \ .$$

This equation can only be true if p or $q = 0$, i.e. the system is either independent of W or deterministic, $p = 1$ or $q = \frac{1}{2}$, which we all excluded by assumption. Hence, $X \not\perp\!\!\!\perp_P W \mid \{Y, Z\}$. The dependence between Z and W given X and Y can be derived in the same way.

S.2 2-ORIENTATION FAITHFULNESS AND SPARSEST MARKOV REPRESENTATION

In this section, we briefly discuss the connection of our new assumptions to approaches based on the sparsest Markov representation (SMR) (Raskutti and Uhler, 2018) which is also referred to as frugality (Forster et al., 2017), which we discussed in the related work section. A graph G^* satisfies the SMR assumption if every graph G that fulfils the Markov property and is not in the Markov equivalence class of G^* contains more edges than G^*. Here we will not discuss the SMR assumption in further detail, but focus on the suggested permutation-based causal discovery algorithm under the SMR assumption, which is called the Sparsest Permutation (SP) algorithm.

To explain the SP algorithm, we need to define a DAG G_π, w.r.t. a permutation π. A DAG G_π consists of vertices $\boldsymbol{V}$ and directed edges E_π, where an edge from the j-th node $\pi(j)$ according to permutation π to node $\pi(k)$ is in E_π if and only if $j < k$ and

$$X_{\pi(j)} \not\perp\!\!\!\perp_P X_{\pi(k)} \mid \{X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(k-1)}\} \backslash \{X_{\pi(j)}\} \ ,$$

where $X_{\pi(j)}$ refers to the j-th random variable according to permutation π. Based on this definition, the SP algorithm constructs a graph G_π for each possible permutation and selects that permutation π^* for which G_{π^*} contains the fewest edges. This permutation π^* is also called minimal or a minimal permutation, if it is not unique.

Although this procedure might be very slow in practice, it has theoretically appealing properties. In particular, we conjecture that it can identify the collider pattern even if strict 2-associations are included, if 2-orientation faithfulness holds. In this work, we will not provide a proof for this conjecture, but give some evidence by discussing the behaviour of the SP algorithm on an example graph.

Consider the graph provided in Figure 4(a) again. For this example, we assume that $\boldsymbol{V}$ does not consist of any further vertices than the four shown in the graph. We will show that all permutations π that are minimal have in common that $\pi(4) = Y$. W.l.o.g. let $\pi(1) = X, \pi(2) = Z$ and $\pi(3) = W$, then G_π only contains the three correct edges, which are:

$$\begin{aligned}
\pi(1) \to \pi(4) &: X \not\perp\!\!\!\perp_P Y \mid \{Z, W\} \\
\pi(2) \to \pi(4) &: Z \not\perp\!\!\!\perp_P Y \mid \{X, W\} \\
\pi(3) \to \pi(4) &: W \not\perp\!\!\!\perp_P Y \mid \{X, Z\}
\end{aligned}$$

and we do not add any superfluous edges, as

$$\begin{aligned}
\pi(1) \to \pi(2) &: X \perp\!\!\!\perp_P Z \mid \emptyset \\
\pi(1) \to \pi(3) &: X \perp\!\!\!\perp_P W \mid Z \\
\pi(2) \to \pi(3) &: Z \perp\!\!\!\perp_P W \mid X \ .
\end{aligned}$$

If we would pick a permutation π' in which we flip for example W and Y such that Y is no longer the node assigned to the highest number in the permutation, i.e. $\pi'(3) = Y$ and $\pi'(4) = W$, we will find more edges and thus not a minimal graph anymore. In particular, we get that

$$\begin{aligned}
\pi'(1) \to \pi'(3) &: X \not\perp\!\!\!\perp_P Y \mid \{Z\} \\
\pi'(2) \to \pi'(3) &: Z \not\perp\!\!\!\perp_P Y \mid \{X\} \\
\pi'(3) \to \pi'(4) &: Y \not\perp\!\!\!\perp_P W \mid \{X, Z\} \\
\pi'(1) \to \pi'(4) &: X \not\perp\!\!\!\perp_P W \mid \{Z, Y\} \\
\pi'(2) \to \pi'(4) &: Z \not\perp\!\!\!\perp_P W \mid \{X, Y\}
\end{aligned}$$

and thus the graph according to this permutation contains two edges more than for permutation π. The main point is that we are now allowed to condition on Y, which opens the paths between X or Z and W. Similarly, assume that we put X as the last node and get the order $\pi'(1) = Z, \pi'(2) = W, \pi'(3) = Y$ and $\pi'(4) = X$, for which

$$\begin{aligned}
\pi'(1) \to \pi'(2) &: Z \perp\!\!\!\perp_P W \mid \emptyset \\
\pi'(1) \to \pi'(3) &: Z \perp\!\!\!\perp_P Y \mid \{W\} \\
\pi'(1) \to \pi'(4) &: Z \not\perp\!\!\!\perp_P X \mid \{W, Y\} \\
\pi'(2) \to \pi'(3) &: W \not\perp\!\!\!\perp_P Y \mid \{Z\} \\
\pi'(2) \to \pi'(4) &: W \not\perp\!\!\!\perp_P X \mid \{Z, Y\} \\
\pi'(3) \to \pi'(4) &: Y \not\perp\!\!\!\perp_P X \mid \{Z, W\}
\end{aligned}$$

and hence, we again find four edges, which is one more than for π. Also, if $\pi'(1) = Y$, we can use it in the conditional to find a dependence between X and Z and at least one dependence between X or Z and W. Hence, the SP algorithm would infer a correct ordering for this graph.

An interesting avenue for future work would be to analyze whether it is possible to always detect the collider pattern also in larger graphs and triples that may be shielded.

S.3 PROOFS

Before we provide the proofs, we state the graphoid axioms (Dawid, 1979; Spohn, 1980; Geiger et al., 1990), which are used in several of our proofs.

Definition 10 (Graphoid Axioms) *Let* $\mathcal{M} = (G, \boldsymbol{V}, P)$, *with* $\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$. *The (semi-)graphoid axioms are the following rules (*$\perp\!\!\!\perp$ *denotes* $\perp\!\!\!\perp_P$ *and* $\perp\!\!\!\perp_G$*)*

1. *Symmetry:* $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z} \Rightarrow \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X} \mid \boldsymbol{Z}$.
2. *Decomposition:* $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z} \Rightarrow \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}$.
3. *Weak Union:* $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z} \Rightarrow \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W} \cup \boldsymbol{Z}$.
4. *Contraction:* $(\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W} \cup \boldsymbol{Z}) \wedge (\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{W} \mid \boldsymbol{Z}) \Rightarrow \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}$.

For separations only on the graph, the graphoid axioms include two additional rules (only for $\perp\!\!\!\perp_G$*).*

5. *Intersection:* $(\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W} \cup \boldsymbol{Z}) \wedge (\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{W} \mid \boldsymbol{Y} \cup \boldsymbol{Z}) \Rightarrow \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}$, *for any pairwise disjoint subsets* $\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$.
6. *Composition:* $(\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}) \wedge (\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{W} \mid \boldsymbol{Z}) \Rightarrow \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}$.

As an illustration why certain rules only hold for graphs and not generally for probability distributions, consider rule (6) and Figure 1(a) again. From the distribution induced by the xor, we find that $Y \perp\!\!\!\perp_P X$ and $Y \perp\!\!\!\perp_P Z$ but we cannot conclude that $Y \perp\!\!\!\perp_P \{X, Z\}$. If, however, in a graph Y is d-separated from X and from Z then Y is d-separated from the set $\{X, Z\}$.

Lemma 1 *Given $\mathcal{M} = (G, \boldsymbol{V}, P)$, let $\{X, Y, Z\} \subseteq \boldsymbol{V}$ form an unfaithful triple in P, then $X \not\perp\!\!\!\perp_P \{Y, Z\}$, $Y \not\perp\!\!\!\perp_P \{X, Z\}$ and $Z \not\perp\!\!\!\perp_P \{X, Y\}$, which in addition implies that $X \not\perp\!\!\!\perp_P Y \mid Z$, $X \not\perp\!\!\!\perp_P Z \mid Y$ and $Y \not\perp\!\!\!\perp_P Z \mid X$.*

Proof: Assume that w.l.o.g. $X \not\perp\!\!\!\perp_P \{Y, Z\}$ is violated. By weak union, we get $X \perp\!\!\!\perp_P Y \mid Z$ which is equivalent to $Y \perp\!\!\!\perp_P X \mid Z$, using symmetry. We know that $Y \perp\!\!\!\perp_P Z$. By contraction, we get that $Y \perp\!\!\!\perp_P \{X, Z\}$. Similarly, we conclude that $Z \perp\!\!\!\perp_P \{X, Y\}$. Altogether, this implies that X, Y, Z would be independent, which is a contradiction.

Each pair of joint dependence and marginal independence, e.g. $X \not\perp\!\!\!\perp_P \{Y, Z\}$ and $X \perp\!\!\!\perp_P Z$, implies a conditional dependence, e.g. $X \not\perp\!\!\!\perp_P Y \mid Z$, by contraction. □

Lemma 2 *Given $\mathcal{M} = (G, \boldsymbol{V}, P)$, let $\{X, Y, Z\} \subseteq \boldsymbol{V}$ form an unfaithful triple in P. If CMC holds, each node in the triple is d-connected to at least one other node in the triple by a path in G.*

Proof: Assume w.l.o.g. that X is *d*-separated from Y and Z in G—i.e. $X \perp\!\!\!\perp_G Y$ and $X \perp\!\!\!\perp_G Z$. By applying the composition axiom, we get that $X \perp\!\!\!\perp_G \{Y, Z\}$. If we apply the causal Markov condition, we get that $X \perp\!\!\!\perp_P \{Y, Z\}$, which is a contradiction to our assumption. □

Theorem 1 *Given $\mathcal{M} = (G, \boldsymbol{V}, P)$ with distinct $X, Y, Z \in \boldsymbol{V}$ and assume that CMC holds. If $\forall \boldsymbol{S} \subseteq \boldsymbol{V} \setminus \{X, Y, Z\}$ it holds that $X \not\perp\!\!\!\perp_P Y \mid Z \cup \boldsymbol{S}$, $X \not\perp\!\!\!\perp_P Z \mid Y \cup \boldsymbol{S}$ and $Y \not\perp\!\!\!\perp_P Z \mid X \cup \boldsymbol{S}$, then one of the three nodes is a collider on a path of length two between the two other nodes, e.g. $X \rightarrow Y \leftarrow Z$ in G.*

Proof: There must be (at least) one node in $\{X, Y, Z\}$ that is not an ancestor of any of the other nodes, say $Z \notin \mathrm{An}(X)$ and $Z \notin \mathrm{An}(Y)$, because of acyclicity. In other words, $X \notin \mathrm{De}(Z)$ and $Y \notin \mathrm{De}(Z)$. The local Markov property states that $Z \perp\!\!\!\perp_G \mathrm{Nd}(Z) \mid \mathrm{Pa}(Z)$ and hence in particular

$$Z \perp\!\!\!\perp_G \{X, Y\} \mid \mathrm{Pa}(Z) .$$

Further, if $|\mathrm{Pa}(Z) \cap \{X, Y\}| < 2$, we get a contradiction with the assumed conditional dependences. Hence $\{X, Y\} \subseteq \mathrm{Pa}(Z)$ and $X \rightarrow Z \leftarrow Y$ is in G. □

Theorem 2 *Assuming that the causal Markov condition holds, the orientation rule in Definition 8 is sound.*

Proof: First, we derive a general statement about the relations between $\boldsymbol{X}$ and $\boldsymbol{Z}$ without further specifying the role of Y. In particular, we show that there always exists a pair $(X, Z) \in \boldsymbol{X} \times \boldsymbol{Z}$ s.t. w.l.o.g.

$$X \perp\!\!\!\perp_G Z \mid \mathrm{Pa}(X) \cup (\boldsymbol{X} \setminus \{X\}) \cup (\boldsymbol{Z} \setminus \{Z\}) , \tag{1}$$

where $\mathrm{Pa}(X) \subseteq \boldsymbol{V} \setminus \boldsymbol{Z}$. Due to acyclicity, there has to exist a node in $\boldsymbol{X} \cup \boldsymbol{Z}$, say X, that is not an ancestor of any node in $(\boldsymbol{X} \cup \boldsymbol{Z}) \setminus \{X\}$ and hence $(\boldsymbol{X} \cup \boldsymbol{Z}) \setminus \{X\} \subseteq \mathrm{Nd}(X)$. By the local Markov condition, we get that $X \perp\!\!\!\perp_G (\boldsymbol{X} \cup \boldsymbol{Z}) \setminus \{X\} \mid \mathrm{Pa}(X)$. Thus, by weak union,

$$X \perp\!\!\!\perp_G Z \mid \mathrm{Pa}(X) \cup (\boldsymbol{X} \setminus \{X\}) \cup (\boldsymbol{Z} \setminus \{Z\}) ,$$

for any $Z \in \boldsymbol{Z}$. Further, $\boldsymbol{Z} \cap \mathrm{Pa}(X) = \emptyset$, as by assumption no pair of nodes $(X, Z) \in \boldsymbol{X} \times \boldsymbol{Z}$ is adjacent in G.

Since $Y \overset{s}{-}_{\leq 2} \boldsymbol{X}$ and $Y \overset{s}{-}_{\leq 2} \boldsymbol{Z}$, we know that Y is at least adjacent to one node in $\boldsymbol{X}$ and one node in $\boldsymbol{Z}$. Hence, Y can take the following roles:

a) Y is a descendent of each node in $\boldsymbol{X} \cup \boldsymbol{Z}$ (which corresponds to $\boldsymbol{X} \rightarrow Y \leftarrow \boldsymbol{Z}$),
b) Y is a non-descendent of each node in $\boldsymbol{X} \cup \boldsymbol{Z}$ and
c) Y is a descendent of at least one node in $\boldsymbol{X} \cup \boldsymbol{Z}$ and a non-descendent of at least one node in $\boldsymbol{X} \cup \boldsymbol{Z}$.

The first statement corresponds to the graph structure implied by rule i) and any possible structure from the latter two is implied by the probabilities found in rule ii). To show these two implications hold, we do a proof by contraposition for each rule.

Hence, to show rule i), we need to prove that if the graph structure is not a collider—i.e. Y takes one of the roles described in b) or c)—then there exists a pair $(X, Z) \in \boldsymbol{X} \times \boldsymbol{Z}$ and there exists a subset $\boldsymbol{S} \subseteq \boldsymbol{V} \setminus \{X, Z\}$ s.t.

$$X \perp\!\!\!\perp_P Z \mid \boldsymbol{S} \cup \{Y\} \cup (\boldsymbol{X} \setminus \{X\}) \cup (\boldsymbol{Z} \setminus \{Z\}) .$$

First, consider all graphs in which Y is a non-descendent of each node in $\boldsymbol{X} \cup \boldsymbol{Z}$ as described in b) We know from statement (1) that, w.l.o.g., there exists a pair $(X, Z) \in \boldsymbol{X} \times \boldsymbol{Z}$ for which $X \perp\!\!\!\perp_G Z \mid \mathrm{Pa}(X) \cup (\boldsymbol{X} \setminus \{X\}) \cup (\boldsymbol{Z} \setminus \{Z\})$. Since $Y \in \mathrm{Nd}(X)$, we will also find that $X \perp\!\!\!\perp_G Z \mid \mathrm{Pa}(X) \cup (\boldsymbol{X} \setminus \{X\}) \cup (\boldsymbol{Z} \setminus \{Z\}) \cup \{Y\}$, where $\mathrm{Pa}(X)$ does not include X or Z. Thus, by CMC we found the required independence. For the cases described in c), again assume that X is not an ancestor of any node in $(\boldsymbol{X} \cup \boldsymbol{Z}) \setminus \{X\}$. To conclude the same statement as previously, we show that X has to be in $\mathrm{De}(Y)$ and thus $Y \in \mathrm{Nd}(X)$. We do this by deriving a contradiction: assume $X \in \mathrm{Nd}(Y)$. If $\boldsymbol{X}$ consists only of the single node X, then X has to be adjacent to Y, $X \in \mathrm{Pa}(Y)$ and hence $X \rightarrow Y$ in G. Thus, Y (and hence X) has to be an ancestor of at least one node in $\boldsymbol{Z}$, by assumption (Y is a non-descendent of at least one node in $\boldsymbol{X} \cup \boldsymbol{Z}$), which is a contradiction. Similarly, if $\boldsymbol{X}$ contains a second node, X', we know by assumption that $X' \in \mathrm{Nd}(X)$. We also know that the triple $\{X, X', Y\}$ has to contain a collider. X cannot be the collider, since $X \notin \mathrm{De}(Y)$ and also X' cannot be the collider since $X \notin \mathrm{An}(X')$. Hence, Y has to be the collider on the path $\langle X, Y, X' \rangle$. As above, at least one node $Z \in \boldsymbol{Z}$ has to be a descendent of Y, by assumption and thus, $X \in \mathrm{An}(Z)$, which is a contradiction.

Last, we prove that the implication in rule ii) holds. Thus, by contraposition, we need to show that if $\boldsymbol{X} \rightarrow Y \leftarrow \boldsymbol{Z}$, then there exists a pair $X, Z \in \boldsymbol{X} \times \boldsymbol{Z}$ s.t. X is conditionally independent of Z given a subset of $\boldsymbol{V} \backslash \{X, Z\}$ that contains $(\boldsymbol{X} \backslash \{X\}) \cup (\boldsymbol{Z} \backslash \{Z\})$ but does not contain Y. From statement (1) there exists a pair $(X, Z) \in \boldsymbol{X} \times \boldsymbol{Z}$ that is d-separated given $\mathrm{Pa}(X) \cup (\boldsymbol{X} \backslash \{X\}) \cup (\boldsymbol{Z} \backslash \{Z\})$. Since Y cannot be in $\mathrm{Pa}(X)$ due to acyclicity, we showed that there exists such a pair of nodes X, Z that can be rendered conditionally independent by a subset of $\boldsymbol{V} \backslash \{X, Z\}$ that contains $(\boldsymbol{X} \backslash \{X\}) \cup (\boldsymbol{Z} \backslash \{Z\})$ but does not contain Y (after applying CMC). □

Corollary 1 *Given $M := (G, \boldsymbol{V}, P)$ with $Y \in \boldsymbol{V}$ and $\boldsymbol{X}, \boldsymbol{Z} \subseteq \boldsymbol{V}$, where $\boldsymbol{X} \cap \boldsymbol{Z} = \emptyset$, $Y \overset{s}{-}_{\leq 2} \boldsymbol{X}$, $Y \overset{s}{-}_{\leq 2} \boldsymbol{Z}$ and no pair of nodes $(X, Z) \in \boldsymbol{X} \times \boldsymbol{Z}$ is adjacent. Assuming that CMC holds, we can detect if condition i) or ii) of 2-orientation faithfulness fails on the triple $\{\boldsymbol{X}, Y, \boldsymbol{Z}\}$.*

Proof: Since we know that $Y \overset{s}{-}_{\leq 2} \boldsymbol{X}$ and $Y \overset{s}{-}_{\leq 2} \boldsymbol{Z}$, we can conclude that, as in the proof of Theorem 2, Y can take three different roles w.r.t. $\boldsymbol{X}$ and $\boldsymbol{Z}$, where role a) corresponds to condition i) in 2-orientation faithfulness and rule i) in the orientation rule and roles b) and c) correspond to condition ii) and rule ii).

Now assume that condition i) in 2-orientation faithfulness fails, that is, the true graph can be described by role a), but there exists a pair $X \in \boldsymbol{X}$ and $Z \in \boldsymbol{Z}$, for which X is independent of Z given a subset of $\boldsymbol{V} \backslash \{X, Z\}$ that contains $Y \cup (\boldsymbol{X} \backslash \{X\}) \cup (\boldsymbol{Z} \backslash \{Z\})$. If this is the case, we cannot apply rule i) of our orientation rule. In addition, we showed in Theorem 2 that for a graph as described by a) rule ii) can never apply. Thus, we can detect this failure of condition i) in 2-orientation faithfulness by noticing that neither rule i) nor ii) of our orientation rule applies.

Next, assume condition ii) in 2-orientation fails. This means that we cannot apply rule ii) of the orientation rule. Again, we showed that for such graphs Y takes either role b) or c), in which case orientation rule i) can never apply. Hence, we can detect if condition ii) in 2-orientation faithfulness fails, since none of the conditions in the orientation rule is met. □

Theorem 3 *Given $M = (G, \boldsymbol{V}, P)$. Assuming that 2-adjacency faithfulness, Assumption 1 and CMC hold, Algorithm 1 correctly identifies $MB(T)$ for $T \in \boldsymbol{V}$.*

Proof: We follow the original correctness proof under the faithfulness assumption (Margaritis and Thrun, 2000), that consists of two main steps. First, we need to show that $\mathrm{MB}(T) \subseteq \boldsymbol{S}$ after the grow phase and second, we need to ensure that all nodes in $\mathrm{MB}(T)$ stay in $\boldsymbol{S}$ during the shrink phase, while nodes not in $\mathrm{MB}(T)$ will be removed from $\boldsymbol{S}$ in the shrink phase.

Grow phase: By assumption (2-adjacency faithfulness), for each node $X \in \mathrm{PC}(T)$, T is either 1-associated to X, or there exists a set $\boldsymbol{X}$ that includes X such that $T \overset{s}{-}_2 \boldsymbol{X}$. If T is 1-associated to a node X, then $T \not\perp\!\!\!\perp_P X \mid \boldsymbol{S}$, if $X \notin \boldsymbol{S}$, hence we will add those nodes. If T is strictly 2-associated to a set $\{X, Z\}$ then $T \not\perp\!\!\!\perp_P X \mid \boldsymbol{S} \cup \{Z\}$ for all $\boldsymbol{S} \subseteq \boldsymbol{V} \backslash \{X, T, Z\}$. Thus, we also add X to $\boldsymbol{S}$, if $X \notin \boldsymbol{S}$ and afterwards also find that $T \not\perp\!\!\!\perp_P Z \mid \boldsymbol{S}$, if $Z \notin \boldsymbol{S}$, since $X \in \boldsymbol{S}$. Hence, all nodes in $\mathrm{PC}(T)$ will be added during the grow phase. Next, we need to consider the spouses of T that do not overlap with $\mathrm{PC}(T)$, hence might not have been added yet.[1] Since we know that eventually $\boldsymbol{S}$ will contain all children of T, we will afterwards also add the corresponding spouses. In particular, we need to consider two classes of spouses S: 1) Spouses that through a child node C are strictly 2-associated to T ($T \overset{s}{-}_2 \{C, S\}$). Those will be added due to the strict 2-association as explained above. 2) Spouses that are not involved in such a strict 2-association. For the latter, we find a conditional dependence between T and S by conditioning on the corresponding child node C (by Assumption 1), which will be in $\boldsymbol{S}$. A special case occurs if a child node C is strictly 2-associated to two spouses S_1 and S_2. Due to Assumption 1, T is dependent on S_1 if we condition on C and S_2, vice versa T is dependent on S_2 if we condition on C and S_1. Similarly to how we add strict 2-associations above, we will also first add one of the two and then the second one. Thus, after the grow phase, $\boldsymbol{S}$ will contain all elements of $\mathrm{MB}(T)$.

Shrink phase: Since it is possible that after the grow phase $\boldsymbol{S}$ is a superset of $\mathrm{MB}(T)$, we need to ensure that in the shrink phase all $W \notin \mathrm{MB}(T)$ will be deleted from $\boldsymbol{S}$ and all $X \in \mathrm{MB}(T)$ will stay in $\boldsymbol{S}$.

First, we show that no node $X \in \mathrm{MB}(T)$ will be removed from $\boldsymbol{S}$. Assume X is the first element in $\mathrm{MB}(T)$ that we attempt to remove from $\boldsymbol{S}$. If $X \in \mathrm{PC}(T)$, by definition of 2-adjacency faithfulness T is either 1-associated to X and hence, X will not be removed, or T is strictly 2-associated to a set $\boldsymbol{X} \subseteq \mathrm{MB}(T)$ that contains X. W.l.o.g. let $\boldsymbol{X} = \{X, Z\}$, then $T \not\perp\!\!\!\perp_P X \mid \boldsymbol{S} \backslash \{X\}$, since $\boldsymbol{S}$ contains Z, and hence, X will not be removed from $\boldsymbol{S}$. If X is a spouse of T, there again exist two cases. Either T is strictly 2-associated to a set that contains X, in which case, X will not be removed from $\boldsymbol{S}$ as explained above, or T is not strictly 2-associated to a set that contains X. In the latter case, by Assumption 1, X is dependent on T conditioned on a subset of $\mathrm{MB}(T) \backslash \{X\}$ and thus $X \not\perp\!\!\!\perp_P T \mid \boldsymbol{S} \backslash \{X\}$. In particular, this subset consists of the common child C and in the special case that C is strictly 2-associated to X and a second spouse S, it also contains that second spouse S. Either way, those conditioning sets are contained in $\boldsymbol{S}$. Hence, X will not be removed from $\boldsymbol{S}$. In the following iterations, $\boldsymbol{S}$ will still contain $\mathrm{MB}(T)$ and hence, we will also not remove a true element of $\mathrm{MB}(T)$.

Last, assume $W \notin \mathrm{MB}(T)$, but $W \in \boldsymbol{S}$ after the grow phase. Further, we can write $\boldsymbol{S} \backslash \{W\}$ as $\mathrm{MB}(T) \cup \boldsymbol{Q}$, where

[1] There could be nodes that are spouses of T and in $\mathrm{PC}(T)$ at the same time e.g. if T has two children X and Z, where Z is also a parent of X.

$\boldsymbol{Q}$ contains all elements from $\boldsymbol{S} \setminus \{W\}$ that are not in MB(T). Then, $T \perp\!\!\!\perp_G \{W\} \cup \boldsymbol{Q} \mid \text{MB}(T)$ and thus by weak union, $T \perp\!\!\!\perp_G W \mid \text{MB}(T) \cup \boldsymbol{Q}$, which implies $T \perp\!\!\!\perp_P W \mid \boldsymbol{S} \setminus \{W\}$ (by CMC). Hence, we delete each node in $\boldsymbol{S}$ that is not in MB(T) in the shrink phase. □

References

A. P. Dawid. Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1):1–15, 1979.

M. Forster, G. Raskutti, R. Stern, and N. Weinberger. The Frugal Inference of Causal Relations. *The British Journal for the Philosophy of Science*, 69(3):821–848, 2017.

D. Geiger, T. Verma, and J. Pearl. Identifying Independence in Bayesian Networks. *Networks*, 20(5):507–534, 1990.

D. Margaritis and S. Thrun. Bayesian Network Induction via Local Neighborhoods. In *Advances in Neural Information Processing Systems*, pages 505–511, 2000.

G. Raskutti and C. Uhler. Learning directed acyclic graph models based on sparsest permutations. *Stat*, 7(1), 2018.

W. Spohn. Stochastic Independence, Causal Independence, and Shieldability. *Journal of Philosophical logic*, 9(1): 73–99, 1980.