Constraint-Based Causal Discovery with Partial Ancestral Graphs in the presence of Cycles

Mooij, J.M.; Claassen, T.

Publication date
2020

Document Version
Final published version

Published in
Proceedings of Machine Learning Research

License
Other

Constraint-Based Causal Discovery using Partial Ancestral Graphs in the presence of Cycles

Joris M. Mooij
Korteweg-de Vries Institute
University of Amsterdam
Amsterdam, The Netherlands

Tom Claassen
Institute for Computing and Information Sciences
Radboud University Nijmegen
Nijmegen, The Netherlands

Abstract

While feedback loops are known to play important roles in many complex systems, their existence is ignored in a large part of the causal discovery literature, as systems are typically assumed to be acyclic from the outset. When applying causal discovery algorithms designed for the acyclic setting on data generated by a system that involves feedback, one would not expect to obtain correct results. In this work, we show that—surprisingly—the output of the Fast Causal Inference (FCI) algorithm is correct if it is applied to observational data generated by a system that involves feedback. More specifically, we prove that for observational data generated by a simple and σ-faithful Structural Causal Model (SCM), FCI is sound and complete, and can be used to consistently estimate (i) the presence and absence of causal relations, (ii) the presence and absence of direct causal relations, (iii) the absence of confounders, and (iv) the absence of specific cycles in the causal graph of the SCM. We extend these results to constraint-based causal discovery algorithms that exploit certain forms of background knowledge, including the causally sufficient setting (e.g., the PC algorithm) and the Joint Causal Inference setting (e.g., the FCI-JCI algorithm).

1 INTRODUCTION

Causal discovery, i.e., establishing the presence or absence of causal relationships between observed variables, is an important activity in many scientific disciplines. Typical approaches to causal discovery from observational data are either score-based, or constraint-based (or a combination of the two). The more generally applicable constraint-based approach, which we focus on in this work, is based on exploiting information in conditional independences in the observed data to draw conclusions about the possible underlying causal structure.

Although many systems of interest in various application domains involve feedback loops or other types of cyclic causal relationships (for example, in economical, biological, chemical, physical, control and climatological systems), most of the existing literature on causal discovery from observational data ignores this and assumes from the outset that the underlying causal system is acyclic. Nonetheless, several algorithms have been developed specifically for the cyclic setting. For example, quite some work has been done for linear systems (e.g., Richardson and Spirtes, 1999; Lacerda et al., 2008; Hyttinen et al., 2010, 2012; Rothenhäusler et al., 2015).

More generally applicable are causal discovery algorithms that exploit conditional independence constraints, without assuming certain restrictions on the parameterizations of the causal models (such as linearity). Pioneering work in this area was done by Richardson (1996b), resulting in the CCD algorithm, the first constraint-based causal discovery algorithm shown to be applicable in a cyclic setting (see also Richardson, 1996a; Richardson and Spirtes, 1999). It was shown to be sound under the assumptions of causal sufficiency, the d-separation Markov property, and d-faithfulness. More recently, other algorithms that are sound under these assumptions (except for the requirement of causal sufficiency) were proposed (Hyttinen et al., 2014; Strobl, 2018).

However, it was already noted by Spirtes (1994, 1995) that the d-separation Markov property assumption can be too strong in general, and he proposed an alternative criterion, making use of the so-called "collapsed graph" construction. More recently, an alternative formulation in terms of the σ-*separation* criterion was introduced, and the corresponding Markov property was shown to

Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), PMLR volume 124, 2020.

hold in a very general setting (Forré and Mooij, 2017). Whereas the Markov property based on σ-separation applies under mild assumptions, the stronger Markov property based on d-separation is limited to more specific settings (e.g., continuous variables with linear relations, or discrete variables, or the acyclic case) (Forré and Mooij, 2017). As discussed in (Forré and Mooij, 2017; Bongers et al., 2020), the σ-separation Markov property seems appropriate for a wide class of cyclic structural causal models with non-linear functional relationships between non-discrete variables, for example structural causal models corresponding to the equilibrium states of dynamical systems governed by random differential equations (Bongers and Mooij, 2018).

Apart from a Markov property, constraint-based causal discovery algorithms need to make some type of faithfulness assumption. A natural extension of the common faithfulness assumption used in the acyclic setting is obtained by replacing d-separation by σ-separation, that we refer to as *σ-faithfulness*. Forré and Mooij (2018) proposed a constraint-based causal discovery algorithm that is sound and complete, assuming the σ-separation Markov property in combination with the σ-faithfulness assumption. However, their algorithm is limited in practice to about 5–7 variables because of the combinatorial explosion in the number of possible causal graphs with increasing number of variables. Interestingly, under the additional assumption of causal sufficiency, the CCD algorithm is also sound under these assumptions (as already noted in Section 4.5 of Richardson, 1996b). Other causal discovery algorithms (LCD (Cooper, 1997), ICP (Peters et al., 2016) and Y-structures (Mani, 2006)), all originally designed for the acyclic setting, have been shown to be sound also in the σ-separation setting (Mooij et al., 2020). The most general scenario (under the additional assumption of causal sufficiency, however) is addressed by the NL-CCD algorithm (Chapter 4 in Richardson, 1996b), which was shown to be sound under the assumptions of the σ-separation Markov property together with the (weaker) d-faithfulness assumption.

One of the classic algorithms for constraint-based causal discovery is the Fast Causal Inference (FCI) algorithm (Spirtes et al., 1995, 1999; Zhang, 2008b). It was designed for the acyclic case, assuming the d-separation Markov property in combination with the d-faithfulness assumption. Recently, it was observed that when run on data generated by cyclic causal models, the accuracy of FCI is actually comparable to its accuracy in the strictly acyclic setting (Figures 25, 26, 29, 31, 32 in Mooij et al., 2020). This is surprising, as it is commonly believed that the application domain of FCI is limited to acyclic causal systems, and one would expect such serious model misspecification to result in glaringly incorrect results.

In this work, we show that when FCI is applied on data from a cyclic causal system that satisfies the σ-separation Markov property and is σ-faithful, its output is still sound and complete. Furthermore, we derive criteria for how to read off various features from the partial ancestral graph output by FCI (specifically, the absence or presence of ancestral relations, direct relations, cyclic relations and confounders). This provides a practical causal discovery algorithm for that setting that is able to handle hundreds or even thousands of variables as long as the underlying causal model is sparse enough, and that is also applicable in the presence of latent confounders. It thus forms a significant improvement over the previous state-of-the-art in causal discovery for the σ-separation setting.

The results we derive in this work are not limited to FCI, but apply to any constraint-based causal discovery algorithm that solves the same task as FCI does, i.e., that estimates the directed partial ancestral graph from conditional independences in the data, e.g., FCI+ (Claassen et al., 2013) and CFCI (Colombo et al., 2012). Our results therefore make constraint-based causal discovery in the presence of cycles as practical as it is in the acyclic case, without requiring any modifications of the algorithms. Our work also provides the first characterization of the σ-Markov equivalence class of directed mixed graphs. We extend our results to variants of algorithms that exploit certain background knowledge, for example, causal sufficiency (e.g., the PC algorithm, Spirtes et al., 2000) or the Joint Causal Inference framework (e.g., the FCI-JCI algorithm, Mooij et al., 2020). For simplicity, we assume no selection bias in this work, but we expect that our results can be extended to allow for that as well.

2 PRELIMINARIES

In Section A (Supplementary Material), we introduce our notation and terminology and provide the reader with a summary of the necessary definitions and results from the graphical causal modeling and discovery literature. For more details, we refer the reader to the literature (Pearl, 2009; Spirtes et al., 2000; Richardson and Spirtes, 2002; Zhang, 2006, 2008b,a; Bongers et al., 2020; Forré and Mooij, 2017). Here, we only give a short high-level overview of the key notions because of space constraints.

There exists a variety of graphical representations of causal models. Most popular are *directed acyclic graphs (DAGs)*, presumably because of their simplicity. DAGs are appropriate under the assumptions of causal sufficiency (i.e., there are no latent common causes of the observed variables), acyclicity (absence of feedback loops) and no selection bias (i.e., there is no implicit conditioning on a common effect of the observed variables). DAGs have many convenient properties, amongst which a Markov property (which has different equivalent formulations, the most prominent one being in terms of the notion of d-separation) and a simple causal interpretation. A more general class of graphs are *acyclic directed mixed graphs (ADMGs)*. These make use of additional bidirected edges to represent latent confounding, and have a similarly convenient Markov property (sometimes referred to as m-separation) and causal interpretation. When also dropping the assumption of acyclicity (thereby allowing for feedback), one can make use of the more general class of *directed mixed graphs (DMGs)*. These graphs can be naturally associated with (possibly cyclic) structural causal models (SCMs) and can represent feedback loops. The corresponding Markov properties and causal interpretation are more subtle (Bongers et al., 2020) than in the acyclic case. Cyclic SCMs can be used, e.g., to describe the causal semantics of the equilibrium states of dynamical systems governed by random differential equations (Bongers and Mooij, 2018).

In this work, we will restrict ourselves to the subclass of *simple* SCMs, i.e., those SCMs for which any subset of the structural equations has a unique solution for the corresponding endogenous variables in terms of the other variables appearing in these equations. Simple SCMs admit (sufficiently weak) cyclic interactions but retain many of the convenient properties of acyclic SCMs (Bongers et al., 2020). They are a special case of modular SCMs (Forré and Mooij, 2017). In particular, they satisfy the σ-separation Markov property and their graphs have an intuitive causal interpretation.[1]

For acyclic constraint-based causal discovery, ADMGs provide a more fine-grained representation than necessary, because one can only recover the Markov equivalence class of ADMGs from conditional independences in observational data. A less expressive class of graphs, *maximal ancestral graphs (MAGs)*, was introduced by Richardson and Spirtes (2002). Each ADMG induces a MAG and each MAG represents a set of ADMGs. The mapping from ADMG to MAG preserves the d-separations and the (non-)ancestral relations. Contrary to ADMGs, MAGs have at most a single edge connecting any pair of distinct variables. One of the key properties that distinguishes MAGs from ADMGs is that Markov-equivalent MAGs have the same adjacencies. In addition to being able to handle latent variables, MAGs can also represent implicit conditioning on a subset of the variables, making use of undirected edges. Therefore, they can be used to represent both latent variables and selection bias.

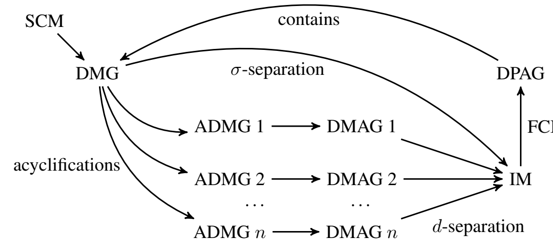

Figure 1: Relations between various representations.

It is often convenient when performing causal reasoning or discovery to be able to represent a set of hypothetical MAGs in a compact way. For these reasons, *partial ancestral graphs (PAGs)* were introduced (Zhang, 2006).[2] The usual way to think about a PAG is as an object that represents a set of MAGs. The (Augmented) Fast Causal Inference (FCI) algorithm (Spirtes et al., 1995, 1999; Zhang, 2008b) takes as input the conditional independences that hold in the data (assumed to be d-Markov and d-faithful w.r.t. a "true" ADMG), and outputs a PAG. As shown in seminal work (Spirtes et al., 1995, 1999; Ali et al., 2005; Zhang, 2008b), the FCI algorithm is sound and complete, and the PAG output by FCI represents the Markov equivalence class of the true ADMG.

In this work, we will for simplicity assume no selection bias. This means that we can restrict ourselves to MAGs without undirected edges, which we refer to as *directed MAGs (DMAGs)*, and PAGs without undirected or circle-tail edges, which we refer to as *directed PAGs* (DPAGs). Almost all proofs will be deferred to Section C (Supplementary Material) because of space constraints.

3 EXTENSIONS TO THE CYCLIC SETTING

The theory of MAGs and PAGs is rather intricate. A natural question is how this theory can be extended when the

[1]The σ-separation criterion is very similar to the d-separation criterion, with the only difference being that σ-separation has as an additional condition for a non-collider to block a path that it has to point to a node in a different strongly connected component. Two nodes in a DMG are said to be in the same strongly connected component if and only if they are both ancestor of each other.

[2]PAGs were originally introduced by Richardson (1996b) in order to represent the output of the CCD algorithm. It was conjectured by Richardson that PAGs could also be used to represent the output of the FCI algorithm, which was originally formulated in terms of Partially Oriented Inducing Path Graphs (POIPGs). This conjecture was proved subsequently by Spirtes. Richardson (p. 102, 1996b) notes: "It is an open question whether or not the set of symbols is sufficiently rich to allow us to represent the class of cyclic graphs with latent variables." In the present work we turned full circle by reinterpreting PAGs as representing properties of DMGs, and have thereby answered this question affirmatively.

assumption of acyclicity is dropped. This does not seem to be straightforward at first sight. An obvious approach would be to generalize the notion of MAGs by adding edge types that represent cycles. However, it would probably require a lot of effort to rederive and reformulate the known results about MAGs and PAGs in this more general setting. In this work, we take another approach: we represent a (possibly cyclic) DMG directly by a DPAG. In order to make this idea precise, we first need to extend the notion of inducing path to the cyclic setting. Our strategy is illustrated in Figure 1.

3.1 INDUCING PATHS

We propose the following generalization of the notion of inducing path (Definition 9) to the σ-separation setting:

Definition 1 *Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F} \rangle$ be directed mixed graph (DMG). An* inducing path (walk) *between two nodes $i, j \in \mathcal{V}$ is a path (walk) in $\mathcal{G}$ between i and j on which every collider is in $\mathrm{AN}_{\mathcal{G}}(\{i, j\})$, and each non-collider on the path (walk), except i and j, only has outgoing directed edges to neighboring nodes on the path (walk) that lie in the same strongly connected component of $\mathcal{G}$.*

This is indeed the proper generalization, since it has the following property.

Proposition 1 *Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F} \rangle$ be a DMG and i, j two distinct vertices in $\mathcal{G}$. Then the following are equivalent:*

(i) There is an inducing path in $\mathcal{G}$ between i and j;
(ii) There is an inducing walk in $\mathcal{G}$ between i and j;
(iii) $i \not\perp^{\sigma}_{\mathcal{G}} j \mid Z$ for all $Z \subseteq \mathcal{V} \setminus \{i, j\}$.

In words: there is an inducing walk (or path) between two nodes in a DMG if and only if the two nodes cannot be σ-separated by any subset of nodes that does not contain either of the two nodes.

3.2 REPRESENTING DMGs BY DPAGs

The following definition forms the key to our approach.

Definition 2 *Let $\mathcal{P}$ be a DPAG and $\mathcal{G}$ a DMG, both with vertex set $\mathcal{V}$. We say that $\mathcal{P}$* contains *$\mathcal{G}$ if all of the following hold:*

(i) two vertices i, j are adjacent in $\mathcal{P}$ if and only if there is an inducing path between i, j in $\mathcal{G}$;
(ii) if $i \,{*\!\!\to}\, j$ in $\mathcal{P}$ (i.e., $i \to j$ in $\mathcal{P}$ or $i \circ\!\!\to j$ in $\mathcal{P}$ or $i \leftrightarrow j$ in $\mathcal{P}$), then $j \notin \mathrm{AN}_{\mathcal{G}}(i)$;
(iii) if $i \to j$ in $\mathcal{P}$ then $i \in \mathrm{AN}_{\mathcal{G}}(j)$.

It is only a slight variation on how PAGs are traditionally interpreted, and agrees with the traditional (acyclic) interpretation when restricting the DMGs to be acyclic.

3.3 ACYCLIFICATIONS

Inspired by the "collapsed graph" construction of Spirtes (1994, 1995), Forré and Mooij (2017) introduced a notion of *acyclification* for a class of graphical causal models termed HEDGes, but the same concept can be defined for DMGs, which we will do here.

Definition 3 *Given a DMG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F} \rangle$. An* acyclification *of $\mathcal{G}$ is an ADMG $\mathcal{G}' = \langle \mathcal{V}, \mathcal{E}', \mathcal{F}' \rangle$ with*

(i) the same nodes $\mathcal{V}$;
(ii) for any pair of nodes $\{i, j\}$ such that $i \notin \mathrm{SC}_{\mathcal{G}}(j)$:
  (a) $i \to j \in \mathcal{E}'$ iff there exists a node k such that $k \in \mathrm{SC}_{\mathcal{G}}(j)$ and $i \to k \in \mathcal{E}$;
  (b) $i \leftrightarrow j \in \mathcal{F}'$ iff there exists a node k such that $k \in \mathrm{SC}_{\mathcal{G}}(j)$ and $i \leftrightarrow k \in \mathcal{F}$;
(iii) for any pair of distinct nodes $\{i, j\}$ such that $i \in \mathrm{SC}_{\mathcal{G}}(j)$: $i \to j \in \mathcal{E}'$ or $i \leftarrow j \in \mathcal{E}'$ or $i \leftrightarrow j \in \mathcal{F}'$.

In words: all strongly connected components are made fully-connected, edges between strongly connected components are preserved, and any edge into a node in a strongly connected component must be copied and made adjacent to all nodes in the strongly connected component. Note that a DMG may have multiple acyclifications. An example is given in Figure 2.

All acyclifications share certain "spurious" edges: the additional incoming directed and adjacent bidirected edges connecting nodes of two different strongly connected components. These have no *causal* interpretation but are necessary to correctly represent the σ-separation properties as d-separation properties. The skeleton of any acyclification $\mathcal{G}'$ of $\mathcal{G}$ equals the skeleton of $\mathcal{G}$ plus additional spurious adjacencies: the edges $i - j$ with $i \,{*\!\!\to}\, k$ and $k \in \mathrm{SC}_{\mathcal{G}}(j)$, and the edges $i - j$ with $i \in \mathrm{SC}_{\mathcal{G}}(j)$ where i and j are not adjacent in $\mathcal{G}$. These "spurious edges" added in any acyclification of a DMG $\mathcal{G}$ correspond with (non-trivial) inducing paths in $\mathcal{G}$.

The "raison d'être" for acyclifications is that they are σ-separation-equivalent to the original DMG, i.e., their σ-independence models agree:

Proposition 2 *For any DMG $\mathcal{G}$ and any acyclification $\mathcal{G}'$ of $\mathcal{G}$, $\mathrm{IM}_{\sigma}(\mathcal{G}) = \mathrm{IM}_{\sigma}(\mathcal{G}') = \mathrm{IM}_{d}(\mathcal{G}')$.*

One particular acyclification that we will make use of repeatedly will be denoted $\mathcal{G}^{\mathrm{acy}}$, and is obtained by replacing all strongly connected components of $\mathcal{G}$ by fully-connected bidirected components without any directed edges (i.e., if $i \in \mathrm{SC}_{\mathcal{G}}(j)$ then $i \leftrightarrow j$ in $\mathcal{G}'$, but neither $i \to j$ nor $j \to i$ in $\mathcal{G}'$). Another useful set of acyclifications is obtained by replacing all strongly connected components of $\mathcal{G}$ by arbitrary fully-connected DAGs, and optionally adding an arbitrary set of bidirected edges.

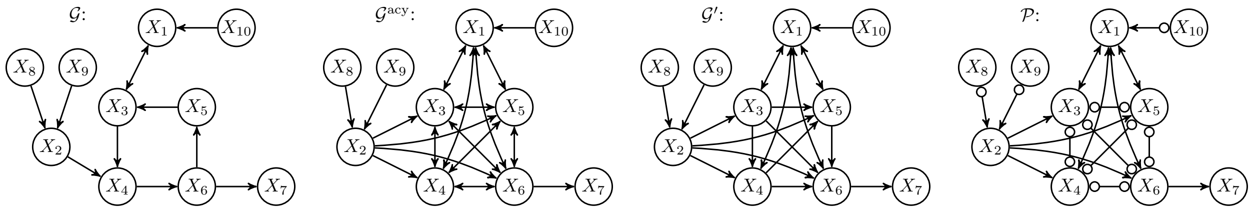

Figure 2: From left to right: Directed mixed graph $\mathcal{G}$, two of its acyclifications ($\mathcal{G}^{\text{acy}}$ and $\mathcal{G}'$) and the DPAG output by FCI $\mathcal{P} = \mathcal{P}_{\texttt{FCI}}(\text{IM}_\sigma(\mathcal{G})) = \mathcal{P}_{\texttt{FCI}}(\text{IM}_d(\mathcal{G}')) = \mathcal{P}_{\texttt{FCI}}(\text{IM}_d(\mathcal{G}^{\text{acy}}))$.

Other important properties of acyclifications are:

Proposition 3 *Let $\mathcal{G}$ be a DMG and i, j two nodes in $\mathcal{G}$.*

(i) If $i \in \text{AN}_{\mathcal{G}}(j)$ then there exists an acyclification $\mathcal{G}'$ of $\mathcal{G}$ with $i \in \text{AN}_{\mathcal{G}'}(j)$;
(ii) If $i \notin \text{AN}_{\mathcal{G}}(j)$ then $i \notin \text{AN}_{\mathcal{G}'}(j)$ for all acyclifications $\mathcal{G}'$ of $\mathcal{G}$;
(iii) There is an inducing path between i and j in $\mathcal{G}$ if and only if there is an inducing path between i and j in $\mathcal{G}'$ for any acyclification $\mathcal{G}'$ of $\mathcal{G}$.

3.4 SOUNDNESS AND COMPLETENESS

In the acyclic setting, the FCI algorithm was shown to be sound and complete (Zhang, 2008b). The notion of acyclifications, together with their elementary properties (Propositions 2 and 3) allows us to easily extend these soundness and completeness results to the σ-separation setting (allowing for cycles).

Consider FCI as a mapping $\mathcal{P}_{\texttt{FCI}}$ from independence models (on variables $\mathcal{V}$) to DPAGs (with vertex set $\mathcal{V}$), which maps the independence model of a DMG $\mathcal{G}$ to the DPAG $\mathcal{P}_{\texttt{FCI}}(\text{IM}_\sigma(\mathcal{G}))$.

Theorem 1 *In the σ-separation setting (but without selection bias), FCI is*

(i) sound*: for all DMGs $\mathcal{G}$, $\mathcal{P}_{FCI}(\text{IM}_\sigma(\mathcal{G}))$ contains $\mathcal{G}$;*
(ii) arrowhead complete*: for all DMGs $\mathcal{G}$: if $i \notin \text{AN}_{\tilde{\mathcal{G}}}(j)$ for any DMG $\tilde{\mathcal{G}}$ that is σ-Markov equivalent to $\mathcal{G}$, then there is an arrowhead $i \leftrightarrow* j$ in $\mathcal{P}_{FCI}(\text{IM}_\sigma(\mathcal{G}))$;*
(iii) tail complete*: for all DMGs $\mathcal{G}$, if $i \in \text{AN}_{\tilde{\mathcal{G}}}(j)$ in any DMG $\tilde{\mathcal{G}}$ that is σ-Markov equivalent to $\mathcal{G}$, then there is a tail $i \to j$ in $\mathcal{P}_{FCI}(\text{IM}_\sigma(\mathcal{G}))$;*
(iv) Markov complete*: for all DMGs $\mathcal{G}_1$ and $\mathcal{G}_2$, $\mathcal{G}_1$ is σ-Markov equivalent to $\mathcal{G}_2$ iff $\mathcal{P}_{FCI}(\text{IM}_\sigma(\mathcal{G}_1)) = \mathcal{P}_{FCI}(\text{IM}_\sigma(\mathcal{G}_2))$.*

Proof sketch: The main idea is the following (see also Figure 1). For all DMGs $\mathcal{G}$, $\text{IM}_\sigma(\mathcal{G}) = \text{IM}_d(\mathcal{G}')$ for any acyclification $\mathcal{G}'$ of $\mathcal{G}$ (Proposition 2). Hence FCI maps any acyclification $\mathcal{G}'$ of $\mathcal{G}$ to the same DPAG $\mathcal{P}_{\texttt{FCI}}(\text{IM}_\sigma(\mathcal{G}))$, and thereby any conclusion we draw about these acyclifications can be transferred back to a conclusion about $\mathcal{G}$ by means of Proposition 3. A complete proof is given in Section C of the Supplementary Material. □

Note that these definitions of soundness and completeness reduce to their acyclic counterparts (Zhang, 2008b) when restricting to ADMGs. In particular, the soundness and Markov completeness properties together imply that the DPAG $\mathcal{P}_{\texttt{FCI}}(\text{IM}_\sigma(\mathcal{G}))$ output by FCI, when given as input the σ-independence model of a DMG $\mathcal{G}$, represents the σ-Markov equivalence class of $\mathcal{G}$. In other words, FCI provides a characterization of the σ-Markov equivalence class of a DMG. This is, to the best of our knowledge, the first such characterization.

In order to read off the independence model from the DPAG $\mathcal{P}_{\texttt{FCI}}(\text{IM}_\sigma(\mathcal{G}))$, one can follow the same procedure as in the acyclic case: first construct a representative DMAG (for details, see Zhang (2008b)) and then apply the d-separation criterion to this DMAG. While the soundness of FCI (Theorem 1(i)) allows us to read off some (non-)ancestral relations from the DPAG output by FCI, this is by far not all causal information that is identifiable from the σ-Markov equivalence class. In the following sections, we will discuss how various causal features can be identified from DPAGs.

3.5 IDENTIFIABLE (NON-)ANCESTRAL RELATIONS

Zhang (2006) conjectured the soundness and completess of a criterion to read off all invariant ancestral relations from a complete DPAG, i.e., to identify the ancestral relations that are present in all Markov equivalent ADMGs that are represented by a complete DPAG. Roumpelaki et al. (2016) proved soundness of the criterion.[3] We extend Theorem 3.1 in (Roumpelaki et al., 2016) to DPAGs and DMGs:

[3] They also claim to have proved completeness, but their proof is flawed: the last part of the proof that aims to prove that u, v are non-adjacent appears to be incomplete.

Proposition 4 *Let $\mathcal{G}$ be a DMG, and let $\mathcal{P}$ be a DPAG that contains $\mathcal{G}$, and such that all unshielded triples in $\mathcal{P}$ have been oriented according to FCI rule $\mathcal{R}0$ (Zhang, 2008b) using $\mathrm{IM}_\sigma(\mathcal{G})$. For two nodes $i \neq j \in \mathcal{P}$: If*

- *there is a directed path from i to j in $\mathcal{P}$, or*
- *there exist uncovered possibly directed paths (see Definition 13) from i to j in $\mathcal{P}$ of the form $i, u, \dots, j$ and $i, v, \dots, j$ such that u, v are distinct non-adjacent nodes in $\mathcal{P}$,*

then $i \in \mathrm{AN}_{\mathcal{G}}(j)$, i.e., i is ancestor of j according to $\mathcal{G}$.

As an example, from the (complete) DPAG in Figure 2 it follows that $X_2 \in \mathrm{AN}_{\mathcal{G}}(X_4)$, and $X_2 \in \mathrm{AN}_{\mathcal{G}}(X_7)$.

Zhang (2006, p. 137) provides a sound and complete criterion to read off definite non-ancestors from a complete DPAG, assuming acyclicity. We can directly extend this criterion to DPAGs and DMGs:

Proposition 5 *Let $\mathcal{G}$ be a DMG, and let $\mathcal{P}$ be a DPAG that contains $\mathcal{G}$. For two nodes $i \neq j \in \mathcal{P}$: if there is no possibly directed path from i to j in $\mathcal{P}$ then $i \notin \mathrm{AN}_{\mathcal{G}}(j)$.*

As an example, from the DPAG in Figure 2 we can read off that X_8 cannot be ancestor of X_1 in $\mathcal{G}$, nor the other way around. However, $X_3 \circ\!\!-\!\!\circ X_6 \rightarrow X_7$ is a possibly directed path in the DPAG, and so X_3 may be (and in this case is) ancestor of X_7 in $\mathcal{G}$.

3.6 IDENTIFIABLE NON-CONFOUNDED PAIRS

While in ADMGs and DMGs confounding is indicated by bidirected edges, in DPAGs confounding can also "hide" behind directed edges. The following notion is of key importance in this regard:

Definition 4 (Zhang (2008a)) *A directed edge $i \rightarrow j$ in a DMAG is said to be* visible *if there is a node k not adjacent to j, such that either there is an edge between k and i that is into i, or there is a collider path between k and i that is into i and every collider on the path is a parent of j. Otherwise $i \rightarrow j$ is said to be* invisible. *The same notion applies to a DPAG, but is then called* definitely visible *(and its negation* possibly invisible*).*

For example, in the DPAG in Figure 2, edge $X_6 \rightarrow X_7$ is definitely visible (by virtue of $X_2 \rightarrow X_6$), as are all edges $X_2 \rightarrow \{X_3, X_4, X_5, X_6\}$ (by virtue of $X_8 \circ\!\!\rightarrow X_2$, or $X_9 \circ\!\!\rightarrow X_2$).

The notion of (in)visibility is closely related with confounding, as shown in Lemma 9 and 10 in Zhang (2008a). To generalize this, we make use of the following Lemma.

Lemma 1 *Let $\mathcal{P}$ be a DPAG that contains DMG $\mathcal{G}$, and let $k \ast\!\!\rightarrow i$ be an edge in $\mathcal{P}$ that is into i. Then there exists an inducing walk in $\mathcal{G}$ between k and i that is into i. If $k \leftrightarrow i$ in $\mathcal{P}$, then there exists an inducing walk in $\mathcal{G}$ between k and i that is both into k and into i.*

This allows us to generalize Lemma 9 in (Zhang, 2008a) to the cyclic setting (with almost identical proof).

Lemma 2 *Let $\mathcal{P}$ be a DPAG, and $i \rightarrow j$ a directed edge in $\mathcal{P}$. If $i \rightarrow j$ is definitely visible in $\mathcal{P}$, then for all DMGs $\mathcal{G}$ contained in $\mathcal{P}$, there exists no inducing walk between i and j in $\mathcal{G}$ that is into i.*

This provides us with a sufficient condition to read off unconfounded pairs of nodes from DPAGs:

Proposition 6 *Let $\mathcal{P}$ be a DMAG and $\mathcal{G}$ be a DMG contained in $\mathcal{P}$. Let $i \neq j$ be two nodes in $\mathcal{P}$. If i and j are not adjacent in $\mathcal{P}$, or if there is a directed edge $i \rightarrow j$ in $\mathcal{P}$ that is definitely visible in $\mathcal{P}$, then $i \leftrightarrow j$ is absent from $\mathcal{G}$.*

For example, from the DPAG in Figure 2 one can infer that there is no bidirected edge $X_2 \leftrightarrow X_7$ in the underlying DMG $\mathcal{G}$, as the two nodes are not adjacent in the DPAG, and also that there is no bidirected edge between X_2 and any node in $\{X_3, X_4, X_5, X_6\}$ in $\mathcal{G}$, as all these edges are definitely visible in the DPAG.

3.7 IDENTIFYING DIRECT (NON-)CAUSES

Contrary to DMGs, a directed edge in a DPAG does not necessarily correspond with a *direct* causal relation. The following proposition provides sufficient conditions to identify the absence of a directed edge from the DPAG.

Proposition 7 *Let $\mathcal{P}$ be a DPAG that contains a DMG $\mathcal{G}$. For two nodes $i \neq j$ in $\mathcal{P}$, if $i \leftarrow\!\!\ast j$ in $\mathcal{P}$, or i and j are not adjacent in $\mathcal{P}$, then $i \rightarrow j$ is not present in $\mathcal{G}$.*

The following proposition was inspired by Theorem 3 in Borboudakis et al. (2012) and provides sufficient conditions to conclude the presence of a directed edge from the DPAG.

Proposition 8 *Let $\mathcal{P}$ be a DPAG that contains a DMG $\mathcal{G}$. For two nodes $i \neq j$ in $\mathcal{P}$, if $i \rightarrow j$ in $\mathcal{P}$ and:*

(i) there does not exist a possibly directed path from i to j in $\mathcal{P}$ that avoids the edge $i \rightarrow j$, or

(ii) if there is no inducing walk between i and j in $\mathcal{G}$ that is both into i and j (for example, because $i \rightarrow j$ is definitely visible in $\mathcal{P}$), and for all vertices k such that there is a possibly directed path $i \ast\!\!-\!\!\ast k \ast\!\!-\!\!\ast j$ from i to j in $\mathcal{P}$, the edge $k \rightarrow j$ is

definitely visible in the DPAG $\mathcal{P}^*$ obtained from $\mathcal{P}$ by replacing the edge between k and j by $k \to j$,

then $i \to j$ is present in $\mathcal{G}$.

As an example, the edge $X_2 \to X_3$ in the DPAG in Figure 2 cannot be identified as being present in $\mathcal{G}$ because both conditions are not satisfied: (i) because of the possibly directed path $X_2 \to X_4 \circ\!\!-\!\!\circ X_3$, (ii) because of the same path where the edge $X_4 \to X_3$ would be possibly invisible if oriented in that way. Also the edge $X_1 \to X_3$ in the DPAG cannot be identified as being present in $\mathcal{G}$. The edge $X_6 \to X_7$ in the DPAG, on the other hand, is identifiably present in $\mathcal{G}$.

3.8 IDENTIFIABLE NON-CYCLES

Strongly connected components in the DMG end up as a specific pattern in the DPAG. This can be used as a sufficient condition for identifying the absence of certain cyclic causal relations in a complete DPAG.

Proposition 9 *Let $\mathcal{G}$ be a DMG and denote by $\mathcal{P} = \mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$ the corresponding complete DPAG output by FCI. Let $i \neq j$ be two nodes in $\mathcal{P}$. If $j \in \mathrm{SC}_{\mathcal{G}}(i)$, then $i \circ\!\!-\!\!\circ j$ in $\mathcal{P}$, and for all nodes k: $k \to i$ in $\mathcal{P}$ iff $k \to j$ in $\mathcal{P}$, and $k \leftrightarrow i$ in $\mathcal{P}$ iff $k \leftrightarrow j$ in $\mathcal{P}$, and $k \circ\!\!\to i$ in $\mathcal{P}$ iff $k \circ\!\!\to j$ in $\mathcal{P}$.*

Hence, any pair of nodes that does not fit this pattern cannot be part of a cycle in $\mathcal{G}$. For example, in the complete DPAG in Figure 2, only the nodes in $\{X_3, X_4, X_5, X_6\}$ might be part of a cycle. For all other pairs of nodes, it follows from Proposition 9 that they cannot be part of a cycle. This sufficient condition is also necessary:

Proposition 10 *Let $\mathcal{G}$ be a DMG and denote by $\mathcal{P} = \mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$ the corresponding complete DPAG output by FCI. Let $i \neq j$ be two nodes in $\mathcal{P}$. If there is an edge $i \circ\!\!-\!\!\circ j$ in $\mathcal{P}$, and all nodes k for which $k \ast\!\!\to i$ is in $\mathcal{P}$ also have an edge of the same type $k \ast\!\!\to j$ (i.e., the two edge marks at k are the same) in $\mathcal{P}$, then there exists a DMG $\tilde{\mathcal{G}}$ with $j \in \mathrm{SC}_{\tilde{\mathcal{G}}}(i)$ that is σ-Markov equivalent to $\mathcal{G}$, but also a DMG $\mathcal{H}$ with $j \notin \mathrm{SC}_{\mathcal{H}}(i)$ that is σ-Markov equivalent to $\mathcal{G}$.*

In other words, under the conditions of this proposition, it is not identifiable from $\mathcal{P}$ alone whether j and i are part of a causal cycle.

4 EXTENSIONS FOR BACKGROUND KNOWLEDGE

In this section, we discuss extensions of our results to situations in which available causal background knowledge is taken into account by causal discovery algorithms. Assume that we have certain background knowledge, formalized as a Boolean function Ψ on the set of all DMGs (indicating for each DMG whether it satisfies the background knowledge). For example, one type of background knowledge commonly considered in the literature (probably mainly for reasons of simplicity) is *causal sufficiency*, which can be formalized by $\Psi(\mathcal{G}) = 1$ iff $\mathcal{G}$ contains no bidirected edges, and $\Psi(\mathcal{G}) = 0$ otherwise. A less trivial example of background knowledge are the JCI Assumptions, which play a central role in the Joint Causal Inference framework (Mooij et al., 2020) for performing causal discovery from multiple datasets that correspond with measurements of a system in different contexts (for example, a combination of observational and different interventional datasets). The latter example will be discussed in more detail in Section 4.3.

4.1 SOUNDNESS AND COMPLETENESS

We first extend the standard notions of soundness and completeness to a setting that involves cycles and background knowledge (but no selection bias).

Definition 5 *Under background knowledge Ψ, a mapping Φ from independence models to DPAGs is called:*

- sound *if for all DMGs $\mathcal{G}$ with $\Psi(\mathcal{G}) = 1$: $\Phi(\mathrm{IM}_\sigma(\mathcal{G}))$ contains $\mathcal{G}$;*
- arrowhead complete *if for all DMGs $\mathcal{G}$ with $\Psi(\mathcal{G}) = 1$: if $i \notin \mathrm{AN}_{\tilde{\mathcal{G}}}(j)$ for any DMG $\tilde{\mathcal{G}}$ with $\Psi(\tilde{\mathcal{G}}) = 1$ that is σ-Markov equivalent to $\mathcal{G}$, then there is an arrowhead $i \leftarrow\!\!\ast\, j$ in $\Phi(\mathrm{IM}_\sigma(\mathcal{G}))$;*
- tail complete *if for all DMGs $\mathcal{G}$ with $\Psi(\mathcal{G}) = 1$: if $i \in \mathrm{AN}_{\tilde{\mathcal{G}}}(j)$ in any DMG $\tilde{\mathcal{G}}$ with $\Psi(\tilde{\mathcal{G}}) = 1$ that is σ-Markov equivalent to $\mathcal{G}$, then there is a tail $i \to j$ in $\Phi(\mathrm{IM}_\sigma(\mathcal{G}))$;*
- Markov complete *if for all DMGs $\mathcal{G}_1, \mathcal{G}_2$ with $\Psi(\mathcal{G}_1) = \Psi(\mathcal{G}_2) = 1$: $\mathcal{G}_1$ is σ-Markov equivalent to $\mathcal{G}_2$ iff $\Phi(\mathrm{IM}_\sigma(\mathcal{G}_1)) = \Phi(\mathrm{IM}_\sigma(\mathcal{G}_2))$.*

It is called complete if it is both arrowhead complete and tail complete.

Note that this reduces to the standard notions (Zhang, 2008b) if $\Psi(\mathcal{G}) = 1$ iff $\mathcal{G}$ is acyclic, while it also reduces to the notions in Theorem 1 if no background knowledge is used (i.e., $\Psi(\mathcal{G}) = 1$ for all $\mathcal{G}$).

We assume that the background knowledge is *compatible with the acyclification* in the following sense:

Assumption 1 *For all DMGs $\mathcal{G}$ with $\Psi(\mathcal{G}) = 1$, the following three conditions hold:*

(i) There exists an acyclification $\mathcal{G}'$ of $\mathcal{G}$ with $\Psi(\mathcal{G}') = 1$;

(ii) For all nodes i, j in $\mathcal{G}$: if $i \in \mathrm{AN}_{\mathcal{G}}(j)$ then there

exists an acyclification $\mathcal{G}'$ of $\mathcal{G}$ with $\Psi(\mathcal{G}') = 1$ such that $i \in \text{AN}_{\mathcal{G}'}(j)$;

(iii) For all nodes i, j in $\mathcal{G}$: if $i \notin \text{AN}_{\mathcal{G}}(j)$ then $i \notin \text{AN}_{\mathcal{G}'}(j)$ for all acyclifications $\mathcal{G}'$ of $\mathcal{G}$ with $\Psi(\mathcal{G}') = 1$.

For example, the background knowledge of "causal sufficiency" satisfies this assumption, as well as the background knowledge of "acyclicity".

The following result is straightforward given all the definitions, but is also quite powerful, as it allows us to directly generalize existing acyclic soundness and completeness results (for certain background knowledge) to the σ-separation setting.

Theorem 2 *Let Ψ be background knowledge that satisfies Assumption 1 and let Φ be a mapping from independence models to DPAGs. Then:*

(i) If Φ is sound for background knowledge Ψ under the additional assumption of acyclicity, then Φ is sound for background knowledge Ψ.

(ii) If Φ is arrowhead (tail) complete for background knowledge Ψ under the additional assumption of acyclicity, then Φ is arrowhead (tail) complete for background knowledge Ψ.

(iii) If Φ is sound and arrowhead complete for background knowledge Ψ under the additional assumption of acyclicity, then Φ is Markov complete.

In the remainder of this section, we will apply this result to two types of background knowledge: causal sufficiency, and the JCI assumptions.

4.2 CAUSAL SUFFICIENCY

We consider the (commonly assumed) background knowledge of "causal sufficiency". This is formalized by $\Psi(\mathcal{G}) = 1$ iff DMG $\mathcal{G}$ contains no bidirected edges. For the acyclic setting, the well-known PC algorithm (Spirtes et al., 2000), adapted with Meek's orientation rules (Meek, 1995a), was shown to be sound and complete. It outputs a so-called Complete Partially Directed Acyclic Graph (CPDAG), which can be interpreted also as a DPAG (by replacing all undirected edges $i - j$ by bicircle edges $i \circ\!\!-\!\!\circ j$). Because this particular background knowledge satisfies Assumption 1, we can apply Theorem 2 to extend the existing acyclic soundness and completeness results to the cyclic setting:

Corollary 1 *The PC algorithm with Meek's orientation rules is sound, arrowhead complete, tail complete and Markov complete (in the σ-separation setting without selection bias).*

We can therefore also apply Propositions 4, 5, to read off the absence or presence of indirect causal relations from the DPAG (obtained from the CPDAG) output by the PC algorithm. Note that the presence or absence of direct causal relations can be easily read off from the DPAG in this case as they are in one-to-one correspondence with directed edges in the DPAG.

4.3 JOINT CAUSAL INFERENCE

Recently, Mooij et al. (2020) proposed FCI-JCI, an extension of FCI that enables causal discovery from data measured in different contexts (for example, if observational data as well as data corresponding to various interventions is available). This is a particular implementation of the general Joint Causal Inference (JCI) framework. For a detailed treatment, we refer the reader to (Mooij et al., 2020); here we only give a brief summary of the JCI assumptions that we need to extend our results on FCI to FCI-JCI.

Definition 6 (JCI Assumptions) *The data-generating mechanism for a system in a context is described by a simple SCM $\mathcal{M}$ with two types of endogenous variables:* system *variables $\{X_i\}_{i\in\mathcal{I}}$ and* context *variables $\{C_k\}_{k\in\mathcal{K}}$. Its graph $\mathcal{G}(\mathcal{M})$ has nodes $\mathcal{I} \cup \mathcal{K}$ (corresponding to system variables and context variables, respectively). The following (optional) JCI Assumptions can be made about the graph $\mathcal{G} := \mathcal{G}(\mathcal{M})$:*

(1) Exogeneity*: No system variable causes any context variable, i.e., $\forall_{k\in\mathcal{K}}\forall_{i\in\mathcal{I}} : i \to k \notin \mathcal{G}$.*

(2) Randomization*: No pair of context and system variable is confounded, i.e., $\forall_{k\in\mathcal{K}}\forall_{i\in\mathcal{I}} : i \leftrightarrow k \notin \mathcal{G}$.*

(3) Genericity*: The induced subgraph $\mathcal{G}(\mathcal{M})_{\mathcal{K}}$ on the context variables is of the following special form: $\forall_{k\neq k'\in\mathcal{K}} : k \leftrightarrow k' \in \mathcal{G} \wedge k \to k' \notin \mathcal{G}$.*

The following Lemma is key to our extensions to the cyclic σ-separation setting.

Lemma 3 *If subset $\{1\}$, $\{1,2\}$, or $\{1,2,3\}$ of the JCI Assumptions holds for a DMG $\mathcal{G}$, then the same subset of assumptions holds for any acyclification of $\mathcal{G}$.*

This trivially implies that these different combinations of the JCI Assumptions satisfy Assumption 1. That allows us to extend the existing acyclic soundness and completeness results for FCI-JCI to the cyclic setting.

FCI-JCI was shown to be sound under the assumption of acyclicity (Theorem 35, Mooij et al., 2020). This gives with Theorem 2:

Corollary 2 *For the background knowledge consisting of JCI Assumptions $\emptyset$, $\{1\}$, $\{1,2\}$ or $\{1,2,3\}$, the corresponding version of FCI-JCI is sound (in the σ-separation setting without selection bias).*

We can therefore also apply Propositions 5 and 6 to read off the absence of indirect causal relations and confounding from the DPAG output by the FCI-JCI algorithm, and Propositions 7 and 8 to read off the absence or presence of direct causal relations. Furthermore, it is clear from its definition that all unshielded triples in the DPAG that FCI-JCI outputs have been oriented according to FCI rule $\mathcal{R}0$. Therefore, we can also apply Proposition 4 to read off the presence of indirect causal relations from the DPAG output by the FCI-JCI algorithm.

Under all three JCI assumptions, stronger results have been derived. In particular, completeness of FCI-JCI has been shown (Theorem 38 Mooij et al., 2020) under the background knowledge of all three JCI Assumptions in the acyclic setting. This gives with Theorem 2:

Corollary 3 *For the background knowledge consisting of JCI Assumptions $\{1, 2, 3\}$, the FCI-JCI algorithm is arrowhead complete, tail complete and Markov complete (in the σ-separation setting without selection bias).*

An important feature of Joint Causal Inference under JCI Assumptions $\{1, 2, 3\}$ is that the direct (non-)targets of interventions need not be known, but can be discovered from the data. The sufficient condition provided in Proposition 42 of Mooij et al. (2020) can be easily generalized to the σ-separation setting as well by observing that under JCI Assumptions $\{1, 2, 3\}$, there cannot be an inducing walk between a system node and a context node that is into both, and then applying Proposition 7 and Proposition 8. For details, see Proposition 12 in Section B of the Supplementary Material.

Furthermore, also Proposition 9 that allows one to identify the absence of cycles can be extended to FCI-JCI under JCI Assumptions $\{1, 2, 3\}$. For details, see Proposition 13 in Section B of the Supplementary Material.

5 DISCUSSION AND CONCLUSION

We have shown that, surprisingly, the FCI algorithm and several of its variants that were designed for the acyclic setting need not be adapted but directly apply also in the cyclic setting under the assumptions of the σ-Markov property, σ-faithfulness, and the absence of selection bias. Furthermore, we have provided sufficient conditions to identify causal features from the DPAG output by FCI and its variants. For convenience, we state this as a corollary, collecting several of our results.

Corollary 4 *Let $\mathcal{M}$ be a simple (possibly cyclic) SCM with graph $\mathcal{G}(\mathcal{M})$ and assume that its distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})$ is σ-faithful w.r.t. the graph $\mathcal{G}(\mathcal{M})$. When using consistent conditional independence tests on an i.i.d. sample of observational data from the induced distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})$ of $\mathcal{M}$, FCI provides a consistent estimate $\hat{\mathcal{P}}$ of the DPAG $\mathcal{P}_{FCI}(\mathrm{IM}_{\sigma}(\mathcal{G}(\mathcal{M})))$ that represents the σ-Markov equivalence class of $\mathcal{G}(\mathcal{M})$. From the estimated DPAG $\hat{\mathcal{P}}$, we obtain consistent estimates for: (i) the absence/presence of (possibly indirect) causal relations according to $\mathcal{M}$ via Propositions 4 and 5; (ii) the absence of confounders according to $\mathcal{M}$ via Proposition 6; (iii) the absence/presence of direct causal relations according to $\mathcal{M}$ via Propositions 7 and 8; (iv) the absence of causal cycles according to $\mathcal{M}$ via Proposition 9.*

A similar conclusion can be formulated for the FCI-JCI algorithm (see Section B of the Supplementary Material). Obviously, our results apply also in the acyclic setting (where σ-separation reduces to d-separation).

One important limitation of the σ-faithfulness assumption is that it excludes the linear and discrete cases. In pioneering work Richardson (1996b) already proposed a constraint-based causal discovery algorithm (NL-CCD) that made use of the σ-separation Markov assumption, while assuming only the d-faithfulness assumption (which is weaker than the σ-faithfulness assumption). In future work, we plan to investigate this setting as well, as well as the possibility of extending our results to a setting that does not rule out selection bias.

Acknowledgements

We are indebted to Jiji Zhang for contributing the proof of Proposition 11. We thank the reviewers for their constructive feedback that helped us improve this paper. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 639466).

References

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

Ali, R. A., Richardson, T. S., Spirtes, P., and Zhang, J. (2005). Towards characterizing Markov equivalence classes for directed acyclic graphs with latent variables. In *Proceedings of the 21th Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 10–17.

Bongers, S., Forré, P., Peters, J., Schölkopf, B., and Mooij, J. M. (2020). Foundations of structural causal models with cycles and latent variables. *arXiv.org preprint*, arXiv:1611.06221v3 [stat.ME].

Bongers, S. and Mooij, J. M. (2018). From random differential equations to structural causal models: the stochastic case. *arXiv.org preprint*, arXiv:1803.08784v2 [cs.AI].

Borboudakis, G., Triantafillou, S., and Tsamardinos, I. (2012). Tools and algorithms for causally interpreting directed edges in maximal ancestral graphs. In *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models (PGM 2012)*, pages 35–42.

Claassen, T., Mooij, J. M., and Heskes, T. (2013). Learning sparse causal models is not NP-hard. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, pages 172–181.

Colombo, D., Maathuis, M. H., Kalisch, M., and Richardson, T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 40(1):294–321.

Cooper, G. F. (1997). A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Mining and Knowledge Discovery*, 1(2):203–224.

Forré, P. and Mooij, J. M. (2017). Markov properties for graphical models with cycles and latent variables. *arXiv.org preprint*, arXiv:1710.08775 [math.ST].

Forré, P. and Mooij, J. M. (2018). Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. In *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Hyttinen, A., Eberhardt, F., and Hoyer, P. (2010). Causal discovery for linear cyclic models with latent variables. In *Proceedings of the Fifth European Workshop on Probabilistic Graphical Models (PGM 2010)*, pages 153–160.

Hyttinen, A., Eberhardt, F., and Hoyer, P. (2012). Learning linear cyclic causal models with latent variables. *Journal of Machine Learning Research*, 13:3387–3439.

Hyttinen, A., Eberhardt, F., and Järvisalo, M. (2014). Constraint-based causal discovery: Conflict resolution with answer set programming. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence, (UAI-14)*, pages 340–349.

Lacerda, G., Spirtes, P., Ramsey, J., and Hoyer, P. O. (2008). Discovering cyclic causal models by independent components analysis. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI-08)*.

Mani, S. (2006). *A Bayesian Local Causal Discovery Framework*. PhD thesis, University of Pittsburg.

Meek, C. (1995a). Causal inference and causal explanation with background knowledge. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 403–411.

Meek, C. (1995b). Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 411–419.

Mooij, J. M., Magliacane, S., and Claassen, T. (2020). Joint causal inference from multiple contexts. *Journal of Machine Learning Research*, 21(99):1–108.

Pearl, J. (1986). A constraint propagation approach to probabilistic reasoning. In *Proceedings of the First Conference on Uncertainty in Artificial Intelligence (UAI-85)*, pages 357–370.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Peters, J., Bühlmann, P., and Meinshausen, N. (2016). Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society, Series B*, 78(5):947–1012.

Richardson, T. (1996a). A discovery algorithm for directed cyclic graphs. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI-96)*.

Richardson, T. and Spirtes, P. (1999). Automated discovery of linear feedback models. In *Computation, Causation, and Discovery*, pages 253–304. MIT Press.

Richardson, T. S. (1996b). *Models of Feedback: Interpretation and Discovery*. PhD thesis, Carnegie-Mellon University.

Richardson, T. S. and Spirtes, P. (2002). Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030.

Rothenhäusler, D., Heinze, C., Peters, J., and Meinshausen, N. (2015). BACKSHIFT: Learning causal cyclic graphs from unknown shift interventions. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 1513–1521.

Roumpelaki, A., Borboudakis, G., Triantafillou, S., and Tsamardinos, I. (2016). Marginal causal consistency in constraint-based causal learning. In *Proceedings of the UAI 2016 Workshop on Causation: Foundation to Application*, number 1792 in CEUR Workshop Proceedings, pages 39–47.

Spirtes, P. (1994). Conditional independence in directed cyclic graphical models for feedback. Technical Report CMU-PHIL-54, Carnegie Mellon University.

Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 499–506.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT press, 2nd edition.

Spirtes, P., Meek, C., and Richardson, T. S. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 499–506.

Spirtes, P., Meek, C., and Richardson, T. S. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In *Computation, Causation and Discovery*, chapter 6, pages 211–252. The MIT Press.

Spirtes, P. and Verma, T. (1992). Equivalence of causal models with latent variables. Technical Report CMU-PHIL-33, Carnegie Mellon University.

Strobl, E. V. (2018). A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias. *International Journal of Data Science and Analytics*, 8:33–56.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20:557–585.

Zhang, J. (2006). *Causal Inference and Reasoning in Causally Insufficient Systems*. PhD thesis, Carnegie Mellon University.

Zhang, J. (2008a). Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474.

Zhang, J. (2008b). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17):1873–1896.

A PRELIMINARIES

In this section of the Supplementary Material, we briefly state all required definitions, notations and results from the literature to make the paper more self-contained.

A.1 GRAPHS

Here we briefly discuss various types of graphs (directed mixed graphs, maximal ancestral graphs, and partial ancestral graphs) and their properties and relationships from the causal discovery literature. For more details, the reader may consult the relevant literature (Spirtes et al., 2000; Richardson and Spirtes, 2002; Zhang, 2006, 2008b,a).

A.1.1 DIRECTED MIXED GRAPHS (DMGs)

A *Directed Mixed Graph* (DMG) is a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F} \rangle$ with nodes $\mathcal{V}$ and two types of edges: *directed* edges $\mathcal{E} \subseteq \{(i,j) : i,j \in \mathcal{V}, i \neq j\}$, and *bidirected* edges $\mathcal{F} \subseteq \{\{i,j\} : i,j \in \mathcal{V}, i \neq j\}$. We will denote a directed edge $(i,j) \in \mathcal{E}$ as $i \to j$ or $j \leftarrow i$, and call i a *parent* of j. We denote all parents of j in the graph $\mathcal{G}$ as $\text{PA}_{\mathcal{G}}(j) := \{i \in \mathcal{V} : i \to j \in \mathcal{E}\}$. We do not allow for self-cycles $i \to i$ here, but multiple edges (at most one of each type, i.e., at most three) between any pair of distinct nodes are allowed. We will denote a bidirected edge $\{i,j\} \in \mathcal{F}$ as $i \leftrightarrow j$ or $j \leftrightarrow i$. Two nodes $i,j \in \mathcal{V}$ are called *adjacent in* $\mathcal{G}$ if if $i \to j \in \mathcal{E}$ or $i \leftarrow j \in \mathcal{E}$ or $i \leftrightarrow j \in \mathcal{F}$.

A *walk* between two nodes $i,j \in \mathcal{V}$ is a tuple $\langle i_0, e_1, i_1, e_2, i_3, \dots, e_n, i_n \rangle$ of alternating nodes and edges in $\mathcal{G}$ ($n \geq 0$), such that all $i_0, \dots, i_n \in \mathcal{V}$, all $e_1, \dots, e_n \in \mathcal{E} \cup \mathcal{F}$, starting with node $i_0 = i$ and ending with node $i_n = j$, and such that for all $k = 1, \dots, n$, the edge e_k connects the two nodes i_{k-1} and i_k in $\mathcal{G}$. If the walk contains each node at most once, it is called a *path*. A *trivial walk (path)* consists just of a single node and zero edges. A *directed walk (path) from* $i \in \mathcal{V}$ *to* $j \in \mathcal{V}$ is a walk (path) between i and j such that every edge e_k on the walk (path) is of the form $i_{k-1} \to i_k$, i.e., every edge is directed and points away from i. By repeatedly taking parents, we obtain the *ancestors* of j: $\text{AN}_{\mathcal{G}}(j) := \{i \in \mathcal{V} : i = i_0 \to i_1 \to \dots \to i_n = j \text{ in } \mathcal{G}\}$. Similarly, we define the *descendants* of i: $\text{DE}_{\mathcal{G}}(i) := \{j \in \mathcal{V} : i = i_0 \to i_1 \to \dots \to i_n = j \text{ in } \mathcal{G}\}$. In particular, each node is ancestor and descendant of itself. A *directed cycle* is a directed path from i to j such that in addition, $j \to i \in \mathcal{E}$. An *almost directed cycle* is a directed path from i to j such that in addition, $j \leftrightarrow i \in \mathcal{F}$. All nodes on directed cycles passing through $i \in \mathcal{V}$ together form the *strongly connected component* $\text{SC}_{\mathcal{G}}(i) := \text{AN}_{\mathcal{G}}(i) \cap \text{DE}_{\mathcal{G}}(i)$ of i. We extend the definitions to sets $I \subseteq \mathcal{V}$ by setting $\text{AN}_{\mathcal{G}}(I) := \cup_{i \in I} \text{AN}_{\mathcal{G}}(i)$, and similarly for $\text{DE}_{\mathcal{G}}(I)$ and $\text{SC}_{\mathcal{G}}(I)$.

A directed mixed graph $\mathcal{G}$ is *acyclic* if it does not contain any directed cycle, in which case it is known as an *Acyclic Directed Mixed Graph (ADMG)*. A directed mixed graph that does not contain bidirected edges is known as a *Directed Graph (DG)*. If a directed mixed graph does not contain bidirected edges and is acyclic, it is called a *Directed Acyclic Graph (DAG)*.

A node i_k on a walk (path) $\pi = \langle i_0, e_1, i_1, e_2, i_3, \dots, e_n, i_n \rangle$ in $\mathcal{G}$ is said to form a *collider on* π if it is a non-endpoint node ($1 \leq k < n$) and the two edges e_k, e_{k+1} meet head-to-head on their shared node i_k (i.e., if the two subsequent edges are of the form $i_{k-1} \to i_k \leftarrow i_{k+1}$, $i_{k-1} \leftrightarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \to i_k \leftrightarrow i_{k+1}$, or $i_{k-1} \leftrightarrow i_k \leftrightarrow i_{k+1}$). Otherwise (that is, if it is an endpoint node, i.e., $k = 0$ or $k = n$, or if the two subsequent edges are of the form $i_{k-1} \to i_k \to i_{k+1}$, $i_{k-1} \leftarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \leftarrow i_k \to i_{k+1}$, $i_{k-1} \leftrightarrow i_k \to i_{k+1}$, or $i_{k-1} \leftarrow i_k \leftrightarrow i_{k+1}$), i_k is called a *non-collider on* π.

The important notion of d-separation was first proposed by Pearl (1986) in the context of DAGs:

Definition 7 (d-separation) *We say that a walk $\langle i_0 \dots i_n \rangle$ in DMG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F} \rangle$ is d-*blocked *by $C \subseteq \mathcal{V}$ if:*

(i) its first node $i_0 \in C$ or its last node $i_n \in C$, or
(ii) it contains a collider $i_k \notin \text{AN}_{\mathcal{G}}(C)$, or
(iii) it contains a non-collider $i_k \in C$.

*If all paths in $\mathcal{G}$ between any node in set $A \subseteq \mathcal{V}$ and any node in set $B \subseteq \mathcal{V}$ are d-blocked by a set $C \subseteq \mathcal{V}$, we say that A is d-*separated *from B by C, and we write $A \perp^{d}_{\mathcal{G}} B \mid C$.*

In the general cyclic case, the notion of d-separation is too strong, as was already pointed out by Spirtes (1995). A solution is to replace it with a non-trivial generalization of d-separation, known as σ-separation:

Definition 8 (σ-separation (Forré and Mooij, 2017)) *We say that a walk $\langle i_0 \dots i_n \rangle$ in DMG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F} \rangle$ is σ-*blocked *by $C \subseteq \mathcal{V}$ if:*

(i) its first node $i_0 \in C$ or its last node $i_n \in C$, or
(ii) it contains a collider $i_k \notin \text{AN}_{\mathcal{G}}(C)$, or
(iii) it contains a non-collider $i_k \in C$ that points to a neighboring node on the walk in another strongly connected component (i.e., $i_{k-1} \to i_k \to i_{k+1}$ or $i_{k-1} \leftrightarrow i_k \to i_{k+1}$ with $i_{k+1} \notin \text{SC}_{\mathcal{G}}(i_k)$, $i_{k-1} \leftarrow i_k \leftarrow i_{k+1}$ or $i_{k-1} \leftarrow i_k \leftrightarrow i_{k+1}$ with $i_{k-1} \notin \text{SC}_{\mathcal{G}}(i_k)$, or $i_{k-1} \leftarrow i_k \to i_{k+1}$ with $i_{k-1} \notin \text{SC}_{\mathcal{G}}(i_k)$ or $i_{k+1} \notin \text{SC}_{\mathcal{G}}(i_k)$).

If all paths in $\mathcal{G}$ between any node in set $A \subseteq \mathcal{V}$ and any node in set $B \subseteq \mathcal{V}$ are σ-blocked by a set $C \subseteq \mathcal{V}$, we say that A is σ-separated *from B by C, and we write $A \perp^{\sigma}_{\mathcal{G}} B \mid C$.*

For a DMG $\mathcal{G}$, define its *d-independence model* to be

$$\mathrm{IM}_d(\mathcal{G}) := \{\langle A, B, C\rangle : A, B, C \subseteq \mathcal{V}, A \underset{\mathcal{G}}{\overset{d}{\perp}} B \mid C\},$$

i.e., the set of all d-separations entailed by the graph, and its *σ-independence model* to be

$$\mathrm{IM}_\sigma(\mathcal{G}) := \{\langle A, B, C\rangle : A, B, C \subseteq \mathcal{V}, A \underset{\mathcal{G}}{\overset{\sigma}{\perp}} B \mid C\},$$

i.e., the set of all σ-separations entailed by the graph. For ADMGs, σ-separation is equivalent to d-separation, and hence, if $\mathcal{G}$ is acyclic, then $\mathrm{IM}_d(\mathcal{G}) = \mathrm{IM}_\sigma(\mathcal{G})$. We call two DMGs $\mathcal{G}_1$ and $\mathcal{G}_2$ *σ-Markov equivalent* if $\mathrm{IM}_\sigma(\mathcal{G}_1) = \mathrm{IM}_\sigma(\mathcal{G}_2)$, and *$d$-Markov equivalent* if $\mathrm{IM}_d(\mathcal{G}_1) = \mathrm{IM}_d(\mathcal{G}_2)$.

A.1.2 DIRECTED MAXIMAL ANCESTRAL GRAPHS (DMAGs)

The following graphical notion will be necessary for the definition of DMAGs.[4]

Definition 9 *Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F}\rangle$ be an acyclic directed mixed graph (ADMG). An* inducing path[5] between two nodes $i, j \in \mathcal{V}$ *is a path in $\mathcal{G}$ between i and j on which every node (except for the endnodes) is a collider on the path and an ancestor in $\mathcal{G}$ of an endnode of the path.*

DMAGs can now be defined as follows (Zhang, 2008a):

Definition 10 *A directed mixed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F}\rangle$ is called a* directed maximal ancestral graph (DMAG) *if all of the following conditions hold:*

1. *Between any two different nodes there is at most one edge, and there are no self-cycles;*
2. *The graph contains no directed or almost directed cycles ("ancestral");*
3. *There is no inducing path between any two non-adjacent nodes ("maximal").*

With the following procedure from Richardson and Spirtes (2002), one can define the DMAG induced by an ADMG:

Definition 11 *Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F}\rangle$ be an ADMG. The* directed maximal ancestral graph induced by $\mathcal{G}$ *is denoted* $\mathrm{DMAG}(\mathcal{G})$ *and is defined as* $\mathrm{DMAG}(\mathcal{G}) = \langle \tilde{\mathcal{V}}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}}\rangle$ *such that $\tilde{\mathcal{V}} = \mathcal{V}$ and for each pair $u, v \in \mathcal{V}$ with $u \neq v$, there is an edge in $\mathrm{DMAG}(\mathcal{G})$ between u and v if and only if there is an inducing path between u and v in $\mathcal{G}$, and in that case the edge in $\mathrm{DMAG}(\mathcal{G})$ connecting u and v is:*

$$\begin{cases} u \to v & \text{if } u \in \mathrm{AN}_{\mathcal{G}}(v), \\ u \leftarrow v & \text{if } v \in \mathrm{AN}_{\mathcal{G}}(u), \\ u \leftrightarrow v & \text{if } u \notin \mathrm{AN}_{\mathcal{G}}(v) \text{ and } v \notin \mathrm{AN}_{\mathcal{G}}(u). \end{cases}$$

This construction preserves the (non-)ancestral relations as well as the d-separations/connections. We sometimes identify a DMAG $\mathcal{H}$ with the set of ADMGs $\mathcal{G}$ that induce $\mathcal{H}$, i.e., such that $\mathrm{DMAG}(\mathcal{G}) = \mathcal{H}$. For a DMAG $\mathcal{H}$, we define its *independence model* to be

$$\mathrm{IM}(\mathcal{H}) := \{\langle A, B, C\rangle : A, B, C \subseteq \mathcal{V}, A \underset{\mathcal{H}}{\overset{d}{\perp}} B \mid C\},$$

i.e., the set of all d-separations entailed by the DMAG. We call two DMAGs $\mathcal{H}_1$ and $\mathcal{H}_2$ *Markov equivalent* if $\mathrm{IM}(\mathcal{H}_1) = \mathrm{IM}(\mathcal{H}_2)$.

A.1.3 DIRECTED PARTIAL ANCESTRAL GRAPHS (DPAGs)

It is often convenient when performing causal reasoning to be able to represent a set of DMAGs in a compact way. For this purpose, *partial ancestral graphs (PAGs)* have been introduced (Zhang, 2006). Again, since we are assuming no selection bias for simplicity, we will only discuss *directed* PAGs (DPAGs), that is, PAGs without undirected or circle-tail edges, i.e., edges of the form $\{ -, -\!\circ, \circ\!- \}$.

Definition 12 *We call a mixed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}\rangle$ with nodes $\mathcal{V}$ and edges $\mathcal{E}$ of the types $\{\to, \leftarrow, \leftarrow\!\circ, \leftrightarrow, \circ\!-\!\circ, \circ\!\!\to\}$ a* directed partial ancestral graph (DPAG) *if all of the following conditions hold:*

1. *Between any two different nodes there is at most one edge, and there are no self-cycles;*
2. *The graph contains no directed or almost directed cycles ("ancestral");*
3. *There is no inducing path between any two non-adjacent nodes ("maximal").*

We extend the definitions of (directed) walks, (directed) paths and colliders for directed mixed graphs to apply also to DPAGs. Edges of the form $i \leftarrow j, i \leftarrow\!\circ\, j, i \leftrightarrow j$ are called *into* i, and similarly, edges of the form $i \to j, i \circ\!\!\to j, i \leftrightarrow j$ are called *into* j. Edges of the form $i \to j$ and $j \leftarrow i$ are called *out of* i.

Given a DMAG or DPAG, its induced *skeleton* is an undirected graph with the same nodes and with an edge be-

[4] We propose an extension of this notion for DMGs in Definition 1.

[5] This was called "primitive inducing path" in (Richardson and Spirtes, 2002).

tween any pair of nodes if and only if the two nodes are adjacent in the DMAG or DPAG.

One often identifies a DPAG with the set of all DMAGs that have the same skeleton as the DPAG, have an arrowhead (tail) on each edge mark for which the DPAG has an arrowhead (tail) at that corresponding edge mark, and for each circle in the DPAG, have either an arrowhead or a tail at the corresponding edge mark. We then say that the DPAG *contains* these DMAGs. Since each ADMG induces a unique DMAG, we can say that a DPAG contains an ADMG if and only if it contains the DMAG induced by it.[6]

We also make use of the following definition (Zhang, 2008b):

Definition 13 *A path $v_0, e_1, v_1, \ldots, v_n$ between nodes v_0 and v_n in a DPAG $\mathcal{P}$ is called a* possibly directed path from v_0 to v_n *if for each $i = 1, \ldots, n$, the edge e_i between v_{i-1} and v_i is not into v_{i-1} (i.e., is of the form $v_{i-1} \circ\!\!-\!\!\circ\, v_i$, $v_{i-1} \circ\!\!\rightarrow v_i$, or $v_{i-1} \rightarrow v_i$). The path is called* uncovered *if every subsequent triple is unshielded, i.e., v_i and v_{i-2} are not adjacent in $\mathcal{P}$ for $i = 2, \ldots, n$.*

A.2 FAST CAUSAL INFERENCE (FCI)

When given as input an independence model $\mathrm{IM}(\mathcal{H})$ of a DMAG $\mathcal{H}$, FCI outputs a DPAG $\mathcal{P}_{\mathrm{FCI}}(\mathrm{IM}(\mathcal{H}))$ that contains $\mathcal{H}$ and is maximally informative (Zhang, 2008b), i.e., each edge mark that is identifiable from the independence model of $\mathcal{H}$ is oriented as such in the DPAG. This DPAG is often referred to as the *complete* DPAG for $\mathcal{H}$.

The following key result seems to be generally known in the field, although we could not easily find a proof in the literature.[7] The proof we provide here is due to Jiji Zhang [private communication].

Proposition 11 *Two MAGs $\mathcal{H}_1, \mathcal{H}_2$ are d-Markov equivalent if and only if $\mathcal{P}_{FCI}(\mathrm{IM}_d(\mathcal{H}_1)) = \mathcal{P}_{FCI}(\mathrm{IM}_d(\mathcal{H}_2))$.*

Proof. We only give a proof for the "if" implication, the "only if" implication being obvious. Ali et al. (2009) showed that two MAGs are Markov equivalent if and only if they have the same skeletons and colliders with order. This implies that colliders with order in any MAG are invariant in the corresponding Markov equivalence class and, by the soundness and arrowhead completeness of FCI, these will appear as colliders in the corresponding PAG. Consider two MAGs $\mathcal{H}_1$ and $\mathcal{H}_2$ that are not Markov equivalent. If they have different skeletons, their corresponding PAGs are not identical. If they do have the same skeletons, then there must be at least one collider with order in $\mathcal{H}_1$ that is not a collider with order in $\mathcal{H}_2$, or vice versa. By Lemma 3.13 in Ali et al. (2009), this would imply that there is at least one collider with order in $\mathcal{H}_1$ that is not a collider in $\mathcal{H}_2$, or vice versa. Hence, because of the soundness and arrowhead completeness of FCI, their corresponding PAGs are not identical. □

Another important property of the FCI algorithm that we use is the following. It was stated in a different but equivalent formulation by Ali et al. (Lemma 4.1, 2005).

Lemma 4 (Lemma A.1 in (Zhang, 2008b)) *Let $\mathcal{H}$ be a MAG and denote by $\mathcal{P} = \mathcal{P}_{FCI}(\mathrm{IM}(\mathcal{H}))$ the corresponding complete PAG output by FCI. Then for any three distinct vertices a, b, c: if $a \ast\!\!\rightarrow b \circ\!\!-\!\!\ast\, c$ in $\mathcal{P}$, then $a \ast\!\!\rightarrow c$ in $\mathcal{P}$; furthermore, if $a \rightarrow b \circ\!\!-\!\!\ast\, c$ in $\mathcal{P}$, then $a \leftrightarrow c$ is not in $\mathcal{P}$.*

Since this property is about arrowhead completeness, it already holds for the DPAG constructed in the first stage of the FCI algorithm, after running the arrowhead orientation rules $\mathcal{R}0$–$\mathcal{R}4$, but before running the tail orientation rules $\mathcal{R}5$–$\mathcal{R}10$.

A.3 STRUCTURAL CAUSAL MODELS (SCMs)

In this subsection we state some of the basic definitions and results regarding Structural Causal Models. Structural Causal Models (SCMs), also known as Structural Equation Models (SEMs), were introduced a century ago by (Wright, 1921) and popularized in AI by Pearl (2009). We follow here the treatment in Bongers et al. (2020) because it deals with cycles in a rigorous way.

Definition 14 *A Structural Causal Model (SCM) is a tuple $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ of:*

(i) a finite index set $\mathcal{I}$ for the endogenous variables in the model;
(ii) a finite index set $\mathcal{J}$ for the latent exogenous variables in the model (disjoint from $\mathcal{I}$);
(iii) a product of standard measurable spaces $\boldsymbol{\mathcal{X}} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$, which define the domains of the endogenous variables;
(iv) a product of standard measurable spaces $\boldsymbol{\mathcal{E}} = \prod_{j \in \mathcal{J}} \mathcal{E}_j$, which define the domains of the exogenous variables;
(v) a measurable function $\boldsymbol{f} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \rightarrow \boldsymbol{\mathcal{X}}$, the causal mechanism*;*
(vi) a product probability measure $\mathbb{P}_{\boldsymbol{\mathcal{E}}} = \prod_{j \in \mathcal{J}} \mathbb{P}_{\mathcal{E}_j}$ on $\boldsymbol{\mathcal{E}}$ specifying the exogenous distribution.

[6] In Definition 2, we propose to extend this to the notion that a DPAG contains a DMG.

[7] For DMAGs and POIPGs, this was already known for an earlier version of FCI; see Corollary 6.4.1 in Spirtes et al. (2000) and the proof in Spirtes and Verma (1992).

Usually, the components of $\boldsymbol{f}$ do not depend on all variables, which is formalized by:

Definition 15 *Let $\mathcal{M}$ be an SCM. We call $i \in \mathcal{I} \cup \mathcal{J}$ a parent of $k \in \mathcal{I}$ if and only if there does not exist a measurable function $\tilde{f}_k : \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{i\}} \times \boldsymbol{\mathcal{E}}_{\mathcal{J}\setminus\{i\}} \to \mathcal{X}_k$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, $x_k = \tilde{f}_k(\boldsymbol{x}_{\mathcal{I}\setminus\{i\}}, \boldsymbol{e}_{\mathcal{J}\setminus\{i\}}) \iff x_k = f_k(\boldsymbol{x}, \boldsymbol{e})$.*

This definition allows us to define the directed mixed graph associated to an SCM (which corresponds with the DMG in Figure 1, our starting point for this work):

Definition 16 *Let $\mathcal{M}$ be an SCM. The induced* graph *of $\mathcal{M}$, denoted $\mathcal{G}(\mathcal{M})$, is defined as the directed mixed graph with nodes $\mathcal{I}$, directed edges $i_1 \to i_2$ iff i_1 is a parent of i_2, and bidirected edges $i_1 \leftrightarrow i_2$ iff there exists $j \in \mathcal{J}$ such that j is parent of both i_1 and i_2.*

If $\mathcal{G}(\mathcal{M})$ is acyclic, we call the SCM $\mathcal{M}$ *acyclic*, otherwise we call the SCM *cyclic*. If $\mathcal{G}(\mathcal{M})$ contains no bidirected edges, we call the endogenous variables in the SCM $\mathcal{M}$ *causally sufficient*.

A pair of random variables $(\boldsymbol{X}, \boldsymbol{E})$ is called a *solution* of the SCM $\mathcal{M}$ if $\boldsymbol{X} = (X_i)_{i\in\mathcal{I}}$ with $X_i \in \mathcal{X}_i$ for all $i \in \mathcal{I}$, $\boldsymbol{E} = (E_j)_{j\in\mathcal{J}}$ with $E_j \in \mathcal{E}_j$ for all $j \in \mathcal{J}$, the distribution $\mathbb{P}(\boldsymbol{E})$ is equal to the exogenous distribution $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$, and the *structural equations*:

$$X_i = f_i(\boldsymbol{X}, \boldsymbol{E}) \quad \text{a.s.}$$

hold for all $i \in \mathcal{I}$.

For acyclic SCMs, solutions exist and have a unique distribution that is determined by the SCM. This is not generally the case in cyclic SCMs, as these could have no solution at all, or could have multiple solutions with different distributions.

Definition 17 *An SCM $\mathcal{M}$ is said to be* uniquely solvable w.r.t. $\mathcal{O} \subseteq \mathcal{I}$ *if there exists a measurable mapping $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{(\text{PA}_{\mathcal{G}(\mathcal{M})}(\mathcal{O})\setminus\mathcal{O})\cap\mathcal{I}} \times \boldsymbol{\mathcal{E}}_{\text{PA}_{\mathcal{G}(\mathcal{M})}(\mathcal{O})\cap\mathcal{J}} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$:*

$$\begin{aligned}\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{(\text{PA}_{\mathcal{G}(\mathcal{M})}(\mathcal{O})\setminus\mathcal{O})\cap\mathcal{I}}, \boldsymbol{e}_{\text{PA}_{\mathcal{G}(\mathcal{M})}(\mathcal{O})\cap\mathcal{J}})\\ \iff \quad \boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e}).\end{aligned}$$

Loosely speaking: the structural equations for $\mathcal{O}$ have a unique solution for $\boldsymbol{X}_{\mathcal{O}}$ in terms of the other variables appearing in those equations. If $\mathcal{M}$ is uniquely solvable with respect to $\mathcal{I}$ (in particular, this holds if $\mathcal{M}$ is acyclic), then it induces a unique *observational distribution* $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})$.

A.3.1 SIMPLE STRUCTURAL CAUSAL MODELS

In this work we restrict attention to a particular subclass of SCMs that has many convenient properties:

Definition 18 *An SCM $\mathcal{M}$ is called* simple *if it is uniquely solvable with respect to each subset $\mathcal{O} \subseteq \mathcal{I}$.*

All acyclic SCMs are simple. The class of simple SCMs can be thought of as a generalization of acyclic SCMs that allows for (sufficiently weak) cyclic causal relations, but preserves many of the convenient properties that acyclic SCMs have. Simple SCMs provide a special case of the more general class of *modular* SCMs (Forré and Mooij, 2017). One of the key aspects of SCMs—which we do not discuss here in detail because we do not make use of it in this work—is their causal semantics, which is defined in terms of *perfect interventions*.

Simple SCMs have the following convenient properties. A simple SCM induces a unique observational distribution. The class of simple SCMs is closed under marginalizations and perfect interventions. Without loss of generality, one can assume that simple SCMs have no self-cycles. The graph of a simple SCM also has a straightforward causal interpretation:

Definition 19 *Let $\mathcal{M}$ be a simple SCM. If $i \to j \in \mathcal{G}(\mathcal{M})$ we call i a* direct cause of j according to $\mathcal{M}$. *If there exists a directed path $i \to \cdots \to j \in \mathcal{G}(\mathcal{M})$, i.e., if $i \in \text{AN}_{\mathcal{G}(\mathcal{M})}(j)$, then we call i a* cause of j according to $\mathcal{M}$. *If there exists a bidirected edge $i \leftrightarrow j \in \mathcal{G}(\mathcal{M})$, then we call i and j* confounded according to $\mathcal{M}$.

The same graph $\mathcal{G}(\mathcal{M})$ of a simple SCM $\mathcal{M}$ also represents the conditional independences that must hold in the observational distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})$ of $\mathcal{M}$. Forré and Mooij (2017) proved a Markov property for the general class of modular SCMs, but we formulate it here only for the special case of simple SCMs:

Theorem 3 (σ-Separation Markov Property) *For any solution $(\boldsymbol{X}, \boldsymbol{E})$ of a simple SCM $\mathcal{M}$, and for all subsets $A, B, C \subseteq \mathcal{I}$ of the endogenous variables:*

$$A \underset{\mathcal{G}(\mathcal{M})}{\overset{\sigma}{\perp}} B \,|\, C \implies \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C.$$

In certain cases, amongst which the acyclic case, the following stronger Markov property holds:

Theorem 4 (d-Separation Markov Property) *For any solution $(\boldsymbol{X}, \boldsymbol{E})$ of an acyclic SCM $\mathcal{M}$, and for all subsets $A, B, C \subseteq \mathcal{I}$ of the endogenous variables:*

$$A \underset{\mathcal{G}(\mathcal{M})}{\overset{d}{\perp}} B \,|\, C \implies \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C.$$

A.3.2 FAITHFULNESS

For a simple SCM $\mathcal{M}$ with endogenous index set $\mathcal{I}$ and observational distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})$, we define its independence model to be

$$\begin{aligned}\mathrm{IM}(\mathcal{M}) := \{\langle A, B, C\rangle : A, B, C \subseteq \mathcal{I},\\ \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{M}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \mid \boldsymbol{X}_C\},\end{aligned}$$

i.e., the set of all (conditional) independences that hold in its (observational) distribution.

The typical starting point for constraint-based approaches to causal discovery from observational data is to assume that the data is modelled by an (unknown) SCM $\mathcal{M}$, such that its observational distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X})$ exists and satisfies a Markov property with respect to its graph $\mathcal{G}(\mathcal{M})$. In other words, $\mathrm{IM}(\mathcal{M}) \supseteq \mathrm{IM}_d(\mathcal{G}(\mathcal{M}))$ in the acyclic case, and more generally, $\mathrm{IM}(\mathcal{M}) \supseteq \mathrm{IM}_\sigma(\mathcal{G}(\mathcal{M}))$ for simple SCMs.

In addition, one usually assumes the *faithfulness assumption* to hold (Spirtes et al., 2000; Pearl, 2009), i.e., that the graph explains *all* conditional independences present in the observational distribution. We distinguish the d-faithfulness assumption:

$$\mathrm{IM}(\mathcal{M}) \subseteq \mathrm{IM}_d(\mathcal{G}(\mathcal{M}))$$

and the σ-faithfulness assumption:[8]

$$\mathrm{IM}(\mathcal{M}) \subseteq \mathrm{IM}_\sigma(\mathcal{G}(\mathcal{M})).$$

Although for d-faithfulness there are some results that this assumption holds *generically* (Meek, 1995b) for certain parameterizations of acyclic SCMs, no such results are known for σ-faithfulness, although it has been conjectured (Spirtes, 1995) that at least weak completeness results can be shown.

B JOINT CAUSAL DISCOVERY

The space constraints forced us to move some material on FCI-JCI into the Supplementary Material, which would have been more appropriate in the main text. We provide these in this section.

Proposition 12 *Let $\mathcal{G}$ be a DMG that satisfies JCI Assumptions $\{1,2,3\}$. Let $\mathcal{P} = \mathcal{P}_{FCI_{JCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$ denote the DPAG output by the corresponding version of FCI-JCI. Let $i \in \mathcal{K}$, $j \in \mathcal{I}$. Then:*

1. *If i is not adjacent to j in $\mathcal{P}$, $i \to j$ is not in $\mathcal{G}$.*
2. *If $i \to j$ in $\mathcal{P}$, and for all system nodes $k \in \mathcal{I}$ s.t. $i \to k$ in $\mathcal{P}$ and k ∘–∘ j or k ∘→ j or $k \to j$ in $\mathcal{P}$, the edge $k \to j$ is definitely visible in the DPAG $\mathcal{P}^*$ obtained from $\mathcal{P}$ by replacing the edge between k and j by $k \to j$, then $i \to j$ is present in $\mathcal{G}$.*

[8]This notion is called *"collapsed graph faithful"* in Richardson (1996b).

Proposition 13 *Let $\mathcal{G}$ be a DMG that satisfies JCI Assumptions $\{1,2,3\}$. Let $\mathcal{P} = \mathcal{P}_{FCI_{JCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$ denote the complete DPAG output by the corresponding version of FCI-JCI. Let $i \neq j$ be two nodes in $\mathcal{P}$. If $j \in \mathrm{SC}_{\mathcal{G}}(i)$, then i ∘–∘ j in $\mathcal{P}$, and for all nodes $k \neq i, j$: $k \to i$ in $\mathcal{P}$ iff $k \to j$ in $\mathcal{P}$, and $k \leftrightarrow i$ in $\mathcal{P}$ iff $k \leftrightarrow j$ in $\mathcal{P}$, and k ∘→ i in $\mathcal{P}$ iff k ∘→ j in $\mathcal{P}$.*

The following corollary collects our results on FCI-JCI.

Corollary 5 *Let $\mathcal{M}$ be a simple (possibly cyclic) SCM with graph $\mathcal{G}(\mathcal{M})$ that satisfies JCI Assumptions $\emptyset$, $\{1\}$, $\{1,2\}$ or $\{1,2,3\}$ and assume that its distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X}, \boldsymbol{C})$ is σ-faithful w.r.t. the graph $\mathcal{G}(\mathcal{M})$. When using consistent conditional independence tests on an i.i.d. sample of data from the induced joint distribution $\mathbb{P}_{\mathcal{M}}(\boldsymbol{X}, \boldsymbol{C})$ of $\mathcal{M}$, FCI-JCI provides a consistent estimate $\hat{\mathcal{P}}$ of the DPAG $\mathcal{P} := \mathcal{P}_{FCI_{JCI}}(\mathrm{IM}_\sigma(\mathcal{G}(\mathcal{M})))$, which contains $\mathcal{G}$. From the estimated DPAG $\hat{\mathcal{P}}$, we obtain consistent esimates for: (i) the absence/presence of (possibly indirect) causal relations according to $\mathcal{M}$ via Propositions 4 and 5; (ii) the absence of confounders according to $\mathcal{M}$ via Proposition 6; (iii) the absence/presence of direct causal relations according to $\mathcal{M}$ via Proposition 8.*

In the special case that $\mathcal{G}(\mathcal{M})$ satisfies all three JCI Assumptions $\{1,2,3\}$, the DPAG $\hat{\mathcal{P}}$ estimated by FCI-JCI also provides consistent estimates for: (iv) the direct intervention targets and non-targets according to $\mathcal{M}$ via Proposition 12; (v) the absence of causal cycles according to $\mathcal{M}$ via Proposition 13. Furthermore, the DPAG $\mathcal{P} := \mathcal{P}_{FCI_{JCI}}(\mathrm{IM}_\sigma(\mathcal{G}(\mathcal{M})))$ characterizes the DMGs that satisfy JCI Assumptions $\{1,2,3\}$ and are σ-Markov equivalent to $\mathcal{G}$.

C PROOFS

In this last section of the Supplementary Material, we provide the proofs for our claims.

Proposition 1 *Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{F}\rangle$ be a DMG and i, j two distinct vertices in $\mathcal{G}$. Then the following are equivalent:*

(i) There is an inducing path in $\mathcal{G}$ between i and j;
(ii) There is an inducing walk in $\mathcal{G}$ between i and j;
(iii) $i \not\perp^\sigma_{\mathcal{G}} j \mid Z$ for all $Z \subseteq \mathcal{V} \setminus \{i, j\}$.

Proof of Proposition 1 The proof is similar to that of Theorem 4.2 in Richardson and Spirtes (2002).

(i) $\implies$ (ii) is trivial.

(ii) $\implies$ (iii): Assume the existence of an inducing walk between i and j in $\mathcal{G}$. Let $Z \subseteq \mathcal{V} \setminus \{i, j\}$. Consider a walk μ in $\mathcal{G}$ between i and j with the fewest number of colliders, with the property that all colliders on μ are in $\mathrm{AN}_{\mathcal{G}}(\{i, j\} \cup Z)$, and each non-endpoint non-collider on μ is not in Z or points only to nodes in the same strongly connected component of $\mathcal{G}$. Such a walk must exist (because the inducing walk satisfies that property). We now show that all colliders on μ must be in $\mathrm{AN}_{\mathcal{G}}(Z)$. Suppose on the contrary the existence of a collider k on μ that is not ancestor of Z. It is either ancestor of i or of j, by assumption. Without loss of generality, assume the latter. Then there is a directed path π from k to j in $\mathcal{G}$ that does not pass through any node of Z. Then the subwalk of μ between i and k can be concatenated with the directed path π into a walk between i and j that has the property, but has fewer colliders than μ: a contradiction. Therefore, μ is σ-open given Z. Hence, i and j are σ-connected given Z.

(iii) $\implies$ (i): Suppose that i and j are σ-connected given any $Z \subseteq \mathcal{V} \setminus \{i, j\}$. In particular, i and j are σ-connected given $Z^* = \mathrm{AN}_{\mathcal{G}}(\{i, j\}) \setminus \{i, j\}$. Let π be a path between i and j that is σ-open given Z^*. We show that π must be an inducing path. First, all colliders on π are in $\mathrm{AN}_{\mathcal{G}}(Z^*)$ and hence in $\mathrm{AN}_{\mathcal{G}}(\{i, j\})$. Second, let k be any non-endpoint non-collider k on π. Then there must be a directed subwalk of π starting at k that ends either at the first collider on π next to k or at an end node of π, and hence k must be in Z^*. Since π is σ-open given Z^*, k can only point to nodes in the same strongly connected component of $\mathcal{G}$. Hence, all non-endpoint non-colliders on π can only point to nodes in the same strongly connected component of $\mathcal{G}$. □

Proposition 2 *For any DMG $\mathcal{G}$ and any acyclification $\mathcal{G}'$ of $\mathcal{G}$, $\mathrm{IM}_\sigma(\mathcal{G}) = \mathrm{IM}_\sigma(\mathcal{G}') = \mathrm{IM}_d(\mathcal{G}')$.*

Proof of Proposition 2 This follows from Theorem 2.8.3 in (Forré and Mooij, 2017). □

Proposition 3 *Let $\mathcal{G}$ be a DMG and i, j two nodes in $\mathcal{G}$.*

(i) If $i \in \mathrm{AN}_{\mathcal{G}}(j)$ then there exists an acyclification $\mathcal{G}'$ of $\mathcal{G}$ with $i \in \mathrm{AN}_{\mathcal{G}'}(j)$;
(ii) If $i \notin \mathrm{AN}_{\mathcal{G}}(j)$ then $i \notin \mathrm{AN}_{\mathcal{G}'}(j)$ for all acyclifications $\mathcal{G}'$ of $\mathcal{G}$;
(iii) There is an inducing path between i and j in $\mathcal{G}$ if and only if there is an inducing path between i and j in $\mathcal{G}'$ for any acyclification $\mathcal{G}'$ of $\mathcal{G}$.

Proof of Proposition 3 (i) If $i \in \mathrm{AN}_{\mathcal{G}}(j)$, then there exists a directed path from i to j in $\mathcal{G}$. Any such directed path visits each strongly connected component of $\mathcal{G}$ at most once. We can choose an acyclification $\mathcal{G}'$ of $\mathcal{G}$ with a suitably chosen DAG on each strongly connected component, in which we can take the shortcut $k \to l$ instead of each longest subpath $k \to u_1 \to \cdots \to u_n \to l$ that consists entirely of nodes within a single strongly connected component of $\mathcal{G}$. This yields a directed path from i to j in $\mathcal{G}'$.

(ii) Let $\mathcal{G}'$ be an acyclification of $\mathcal{G}$. Each directed edge $k \to l$ in $\mathcal{G}'$ is either in $\mathcal{G}$ or corresponds with $k \in \mathrm{AN}_{\mathcal{G}}(l)$. Hence all ancestral relations in $\mathcal{G}'$ must be present in $\mathcal{G}$ as well.

(iii) Follows directly from the separation properties of inducing paths and Proposition 2. □

Theorem 1 *In the σ-separation setting (but without selection bias), FCI is*

(i) sound*: for all DMGs $\mathcal{G}$, $\mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$ contains $\mathcal{G}$;*
(ii) arrowhead complete*: for all DMGs $\mathcal{G}$: if $i \notin \mathrm{AN}_{\tilde{\mathcal{G}}}(j)$ for any DMG $\tilde{\mathcal{G}}$ that is σ-Markov equivalent to $\mathcal{G}$, then there is an arrowhead $i \leftarrow\!\ast\ j$ in $\mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$;*
(iii) tail complete*: for all DMGs $\mathcal{G}$, if $i \in \mathrm{AN}_{\tilde{\mathcal{G}}}(j)$ in any DMG $\tilde{\mathcal{G}}$ that is σ-Markov equivalent to $\mathcal{G}$, then there is a tail $i \to j$ in $\mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$;*
(iv) Markov complete*: for all DMGs $\mathcal{G}_1$ and $\mathcal{G}_2$, $\mathcal{G}_1$ is σ-Markov equivalent to $\mathcal{G}_2$ iff $\mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}_1)) = \mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}_2))$.*

Proof of Theorem 1 To prove soundness, let $\mathcal{G}$ be a DMG and let $\mathcal{P} = \mathcal{P}_{\mathrm{FCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$. The acyclic soundness of FCI means that for all ADMGs $\mathcal{G}'$, $\mathcal{P}_{\mathrm{FCI}}(\mathrm{IM}_d(\mathcal{G}'))$ contains $\mathcal{G}'$. Hence, by Proposition 2, $\mathcal{P}$ contains $\mathcal{G}'$ for all acyclifications $\mathcal{G}'$ of $\mathcal{G}$. But then $\mathcal{P}$ must contain $\mathcal{G}$:

- if two vertices i, j are adjacent in $\mathcal{P}$ then there is an inducing path between i, j in any acyclification of $\mathcal{G}$, which holds if and only if there is an inducing path between i, j in $\mathcal{G}$ (Proposition 3(iii);
- if $i \ast\!\!\to j$ in $\mathcal{P}$, then $j \notin \mathrm{AN}_{\mathcal{G}'}(i)$ for any acyclification $\mathcal{G}'$ of $\mathcal{G}$, and hence $j \notin \mathrm{AN}_{\mathcal{G}}(i)$ (Proposition 3(i));
- if $i \to j$ in $\mathcal{P}$, then $i \in \mathrm{AN}_{\mathcal{G}'}(j)$ for all acyclifications $\mathcal{G}'$ of $\mathcal{G}$, and hence $i \in \mathrm{AN}_{\mathcal{G}}(j)$ (Proposition 3(ii)).

To prove arrowhead completeness, let $\mathcal{G}$ be a DMG and suppose that $i \in \mathrm{AN}_{\tilde{\mathcal{G}}}(j)$ in any DMG $\tilde{\mathcal{G}}$ that is σ-Markov equivalent to $\mathcal{G}$. Since $\mathcal{G}^{\mathrm{acy}}$ is σ-Markov equivalent to $\mathcal{G}$, this implies in particular that for all ADMGs $\tilde{\mathcal{G}}$ that are d-Markov equivalent to $\mathcal{G}^{\mathrm{acy}}$, $i \in \mathrm{AN}_{\tilde{\mathcal{G}}}(j)$. Because of the acyclic arrowhead completeness of FCI, there must be an arrowhead $i \leftarrow\!* \; j$ in $\mathcal{P}_{\texttt{FCI}}(\mathrm{IM}_d(\mathcal{G}^{\mathrm{acy}})) = \mathcal{P}_{\texttt{FCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$. Tail completeness is proved similarly.

To prove Markov completeness: Proposition 2 implies both $\mathrm{IM}_\sigma(\mathcal{G}_1) = \mathrm{IM}_d(\mathcal{G}_1^{\mathrm{acy}})$ and $\mathrm{IM}_\sigma(\mathcal{G}_2) = \mathrm{IM}_d(\mathcal{G}_2^{\mathrm{acy}})$. From the acyclic Markov completeness of FCI (Proposition 11 in the Supplementary Material), it then follows that $\mathcal{G}_1^{\mathrm{acy}}$ must be d-Markov equivalent to $\mathcal{G}_2^{\mathrm{acy}}$, and hence $\mathcal{G}_1$ must be σ-Markov equivalent to $\mathcal{G}_2$.

Alternatively, the statement of this theorem can be seen to be a special case of Theorem 2, applied with the trivial background knowledge $\Psi(\mathcal{G}) = 1$ for all DMGs $\mathcal{G}$, to $\Phi : \mathcal{G} \mapsto \mathcal{P}_{\texttt{FCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$, combined with the known (acyclic) soundness and completeness results of FCI (Zhang, 2008b). Note here that the trivial background knowledge $\Psi(\mathcal{G}) = 1$ satisfies Assumption 1 as follows immediately from Proposition 3. $\square$

Proposition 4 *Let $\mathcal{G}$ be a DMG, and let $\mathcal{P}$ be a DPAG that contains $\mathcal{G}$, and such that all unshielded triples in $\mathcal{P}$ have been oriented according to FCI rule $\mathcal{R}0$ (Zhang, 2008b) using $\mathrm{IM}_\sigma(\mathcal{G})$. For two nodes $i \neq j \in \mathcal{P}$: If*

- *there is a directed path from i to j in $\mathcal{P}$, or*
- *there exist uncovered possibly directed paths (see Definition 13) from i to j in $\mathcal{P}$ of the form $i, u, \dots, j$ and $i, v, \dots, j$ such that u, v are distinct non-adjacent nodes in $\mathcal{P}$,*

then $i \in \mathrm{AN}_\mathcal{G}(j)$, i.e., i is ancestor of j according to $\mathcal{G}$.

Proof of Proposition 4 If there is a directed path from i to j in $\mathcal{P}$, say $v_1 \to \cdots \to v_n$ with $v_1 = i$ and $v_n = j$, then $v_m \in \mathrm{AN}_\mathcal{G}(v_{m+1})$ for all $m = 1, \dots, n-1$. Hence $i \in \mathrm{AN}_\mathcal{G}(j)$.

Otherwise, assume that there exist uncovered possibly directed paths from i to j in $\mathcal{P}$ of the form $i, u, \dots, j$ and $i, v, \dots, j$ such that u, v are distinct and non-adjacent in $\mathcal{P}$. If $i \in \mathrm{AN}_\mathcal{G}(u)$, the path $i, u, \dots, j$ must actually correspond with a directed path from i to j in $\mathcal{G}$, because otherwise it would contain unshielded colliders that were not oriented, contradicting the assumptions. If $i \notin \mathrm{AN}_\mathcal{G}(u)$ instead, one obtains that $i \in \mathrm{AN}_\mathcal{G}(v)$ to avoid an unshielded collider $u \ast\!\!\to i \leftarrow\!\!\ast\; v$ in $\mathcal{P}$ that was not oriented. Hence the path $i, v, \dots, j$ must correspond with a directed path from i to j in $\mathcal{G}$, because otherwise it would contain unshielded colliders that were not oriented, contradicting the assumptions. In both cases, $i \in \mathrm{AN}_\mathcal{G}(j)$. $\square$

Proposition 5 *Let $\mathcal{G}$ be a DMG, and let $\mathcal{P}$ be a DPAG that contains $\mathcal{G}$. For two nodes $i \neq j \in \mathcal{P}$: if there is no possibly directed path from i to j in $\mathcal{P}$ then $i \notin \mathrm{AN}_\mathcal{G}(j)$.*

Proof of Proposition 5 If $i \in \mathrm{AN}_\mathcal{G}(j)$, then the directed path from i to j must correspond with a possibly directed path in $\mathcal{P}$. $\square$

Lemma 1 *Let $\mathcal{P}$ be a DPAG that contains DMG $\mathcal{G}$, and let $k \ast\!\!\to i$ be an edge in $\mathcal{P}$ that is into i. Then there exists an inducing walk in $\mathcal{G}$ between k and i that is into i. If $k \leftrightarrow i$ in $\mathcal{P}$, then there exists an inducing walk in $\mathcal{G}$ between k and i that is both into k and into i.*

Proof of Lemma 1 If $k \ast\!\!\to i$ in $\mathcal{P}$, then there exists an inducing walk between k and i in $\mathcal{G}$ because k and i are adjacent in $\mathcal{P}$ and $\mathcal{P}$ contains $\mathcal{G}$. If this inducing walk were out of i, it would be of the form $k \dots \ast\!\!\to u_n \leftarrow u_{n-1} \leftarrow \cdots \leftarrow u_1 \leftarrow i$, where u_n is the first collider on the walk that one encounters when following the directed edges out of i. u_n must be ancestor of i or k in $\mathcal{G}$, and it cannot be ancestor of k (because then i would be ancestor of k, contradicting the orientation $k \ast\!\!\to i$ in $\mathcal{P}$), hence it must be ancestor of i. Thus there exists a walk $k \dots \ast\!\!\to u_n \to \cdots \to i$ in $\mathcal{G}$ where we replaced the subwalk $u_n \leftarrow u_{n-1} \leftarrow \cdots \leftarrow i$ by a directed path from u_n to i. It is clear that this is an inducing path in $\mathcal{G}$ between k and i that is into i.

If $k \leftrightarrow i$ in $\mathcal{P}$, then by similar reasoning, we obtain an inducing path in $\mathcal{G}$ between k and i that is into k as well as into i. $\square$

Lemma 2 *Let $\mathcal{P}$ be a DPAG, and $i \to j$ a directed edge in $\mathcal{P}$. If $i \to j$ is definitely visible in $\mathcal{P}$, then for all DMGs $\mathcal{G}$ contained in $\mathcal{P}$, there exists no inducing walk between i and j in $\mathcal{G}$ that is into i.*

Proof of Lemma 2 Suppose $\mathcal{G}$ is a DMG contained in $\mathcal{P}$. Then $i \in \mathrm{AN}_\mathcal{G}(j)$ and there exists an inducing walk between i and j in $\mathcal{G}$. We will prove the contrapositive. Assume that there exists an inducing walk in $\mathcal{G}$ between i and j that is into i. Let k be another vertex in $\mathcal{P}$.

If $k \ast\!\!\to i$ in $\mathcal{P}$, then there is an inducing walk between k and i in $\mathcal{G}$ that is into i by Lemma 1.

If there is a collider path π in $\mathcal{P}$ from k to i that is into i and such that every non-endpoint vertex on the walk is parent of j in $\mathcal{P}$, then there is an inducing walk between k and i in $\mathcal{G}$ that is into i. For each pair of adjacent vertices (v_i, v_{i+1}) on π, Lemma 1 gives the existence of an inducing walk in $\mathcal{G}$ between v_i and v_{i+1} that is into v_{i+1}, and also into v_i unless possibly $v_i = k$. Because each vertex other than k on π is ancestor in $\mathcal{G}$ of j, all these inducing walks can be concatenated into one inducing walk in $\mathcal{G}$ between k and i that is into i.

Concatenating the inducing walk between k and i that is into i with the inducing walk between i and j that is into

i we obtain an inducing walk between k and j (note that i becomes a collider that is ancestor of j) in $\mathcal{G}$. Hence k and j are adjacent in $\mathcal{P}$. Since this holds for all k, the directed edge $i \to j$ in $\mathcal{P}$ cannot be definitely visible. $\square$

Proposition 6 *Let $\mathcal{P}$ be a DMAG and $\mathcal{G}$ be a DMG contained in $\mathcal{P}$. Let $i \neq j$ be two nodes in $\mathcal{P}$. If i and j are not adjacent in $\mathcal{P}$, or if there is a directed edge $i \to j$ in $\mathcal{P}$ that is definitely visible in $\mathcal{P}$, then $i \leftrightarrow j$ is absent from $\mathcal{G}$.*

Proof of Proposition 6 If i and j are not adjacent in $\mathcal{P}$, then there is no inducing path between i and j in $\mathcal{G}$ by assumption, and in particular, this rules out the presence of the bidirected edge $i \leftrightarrow j$ in $\mathcal{G}$.

If $i \to j$ in $\mathcal{P}$ is definitely visible, by Lemma 2, there cannot be a bidirected edge $i \leftrightarrow j$ in any DMG contained in $\mathcal{P}$. $\square$

Proposition 7 *Let $\mathcal{P}$ be a DPAG that contains a DMG $\mathcal{G}$. For two nodes $i \neq j$ in $\mathcal{P}$, if $i \leftarrow\!\!* \, j$ in $\mathcal{P}$, or i and j are not adjacent in $\mathcal{P}$, then $i \to j$ is not present in $\mathcal{G}$.*

Proof of Proposition 7 If $i \leftarrow\!\!* \, j$ in $\mathcal{P}$, then $i \notin \mathrm{AN}_{\mathcal{G}}(j)$ and hence $i \to j$ cannot be present in $\mathcal{G}$. If i and j are not adjacent in $\mathcal{P}$, then $i \to j$ cannot be present in $\mathcal{G}$ because this would be an inducing path between i and j. $\square$

Proposition 8 *Let $\mathcal{P}$ be a DPAG that contains a DMG $\mathcal{G}$. For two nodes $i \neq j$ in $\mathcal{P}$, if $i \to j$ in $\mathcal{P}$ and:*

(i) there does not exist a possibly directed path from i to j in $\mathcal{P}$ that avoids the edge $i \to j$, or

(ii) if there is no inducing walk between i and j in $\mathcal{G}$ that is both into i and j (for example, because $i \to j$ is definitely visible in $\mathcal{P}$), and for all vertices k such that there is a possibly directed path $i *\!\!\rightarrow\!\!* \, k *\!\!\rightarrow\!\!* \, j$ from i to j in $\mathcal{P}$, the edge $k \to j$ is definitely visible in the DPAG $\mathcal{P}^*$ obtained from $\mathcal{P}$ by replacing the edge between k and j by $k \to j$,

then $i \to j$ is present in $\mathcal{G}$.

Proof of Proposition 8 (i) Suppose $i \to j$ in $\mathcal{P}$. We prove the contrapositive. Assume $i \to j$ is absent from $\mathcal{G}$. Because $i \in \mathrm{AN}_{\mathcal{G}}(j)$ by assumption, there must be a directed path from i to j in $\mathcal{G}$ that does not contain the edge $i \to j$. This corresponds with a possibly directed path in $\mathcal{P}$ that avoids the edge $i \to j$.

(ii) Suppose $i \to j$ in $\mathcal{P}$. There must be an inducing walk π between i and j in $\mathcal{G}$ that is into j by Lemma 1. Each collider on π is ancestor of j (because it is ancestor of i or j by definition, and i is ancestor of j). By assumption (or by Lemma 2 if $i \to j$ is definitely visible in $\mathcal{P}$), this inducing walk must be out of i.

We now show that π cannot contain any colliders under the assumptions made. For the sake of contradiction, assume that π contained one or more colliders. Denote the collider closest to i on π by k. Since π is out of i and k is the collider on π closest to i, π must start as a directed walk $i \to \cdots \to k$. This is an inducing walk between i and k, and since $i \in \mathrm{AN}_{\mathcal{G}}(k)$, it corresponds with a possibly directed path $i *\!\!\rightarrow\!\!* \, k$ in $\mathcal{P}$. The subwalk of π between k and j is an inducing walk in $\mathcal{G}$ between k and j that is both into k and into j, since each collider on it is ancestor of j and each non-endpoint non-collider only points to nodes in the same strongly connected component. Since $k \in \mathrm{AN}_{\mathcal{G}}(j)$, this corresponds with a possibly directed path $k *\!\!\rightarrow\!\!* \, j$ in $\mathcal{P}$. It also implies that $\mathcal{P}^*$, obtained from $\mathcal{P}$ by replacing the edge between k and j by the directed edge $k \to j$, contains $\mathcal{G}$. By Lemma 2 applied to $\mathcal{P}^*$, $k \to j$ cannot be definitely visible in $\mathcal{P}^*$. We have arrived at a contradiction with the assumption.

Hence, the inducing walk π cannot contain any colliders. If it consisted of multiple edges, it would be of the form $i \to k' \to \cdots \to j$, where now $k' \neq j$ is the vertex on π next to i, and all non-endpoint noncolliders would point to nodes in the same strongly connected component. Hence k' and j would lie in the same strongly connected component of $\mathcal{G}$. Again, note that this results in a possibly directed path $i *\!\!\rightarrow\!\!* \, k' *\!\!\rightarrow\!\!* \, j$ in $\mathcal{P}$, and means that $\mathcal{P}^*$ obtained from $\mathcal{P}$ by replacing the edge between k' and j by the directed edge $\to j$, contains $\mathcal{G}$. By Lemma 2 applied to $\mathcal{P}^*$, there exists no inducing walk in $\mathcal{G}$ between k' and j that is into k', because $k' \to j$ must be definitely visible in $\mathcal{P}^*$ by assumption. This contradicts the existence of a directed path $k' \leftarrow \cdots \leftarrow j$ in $\mathcal{G}$.

Hence, the inducing walk π in $\mathcal{G}$ must consist of a single edge, and is necessarily of the form $i \to j$. $\square$

Proposition 9 *Let $\mathcal{G}$ be a DMG and denote by $\mathcal{P} = \mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$ the corresponding complete DPAG output by FCI. Let $i \neq j$ be two nodes in $\mathcal{P}$. If $j \in \mathrm{SC}_{\mathcal{G}}(i)$, then $i \circ\!\!-\!\!\circ \, j$ in $\mathcal{P}$, and for all nodes k: $k \to i$ in $\mathcal{P}$ iff $k \to j$ in $\mathcal{P}$, and $k \leftrightarrow i$ in $\mathcal{P}$ iff $k \leftrightarrow j$ in $\mathcal{P}$, and $k \circ\!\!\rightarrow i$ in $\mathcal{P}$ iff $k \circ\!\!\rightarrow j$ in $\mathcal{P}$.*

Proof of Proposition 9 Since no pair of nodes within a strongly connected component of $\mathcal{G}$ can be σ-separated, each strongly connected component of $\mathcal{G}$ ends up as a fully-connected component in $\mathcal{P}$. For two nodes $i \neq j$ in the same strongly connected component of $\mathcal{G}$, there exists an acyclification of $\mathcal{G}$ in which $i \to j$ and another one in which $i \leftarrow j$, and hence the edge between i and j in $\mathcal{P}$ must be oriented as $i \circ\!\!-\!\!\circ \, j$. From Lemma 4 it then directly follows that for any third node k, $k *\!\!\rightarrow i \in \mathcal{P}$ if and only if $k *\!\!\rightarrow j \in \mathcal{P}$. If $k \leftrightarrow i \in \mathcal{P}$, then $k \circ\!\!\rightarrow j \notin \mathcal{P}$, otherwise $i \leftrightarrow k \circ\!\!\rightarrow j$ would violate Lemma 4. Also, if $k \leftrightarrow i \in \mathcal{P}$, then $k \rightarrow j \notin \mathcal{P}$, otherwise $k \rightarrow j \circ\!\!-\!\!\circ i$ would violate Lemma 4.

Hence, we have shown that for all $i \neq j$ with $i \in \mathrm{SC}_{\mathcal{G}}(j)$, $i \circ\!\!-\!\!\circ j \in \mathcal{P}$ and for all k:

$$\begin{cases} k \ast\!\!\rightarrow i \in \mathcal{P} \iff k \ast\!\!\rightarrow j \in \mathcal{P}, \text{ and} \\ k \leftrightarrow i \in \mathcal{P} \iff k \leftrightarrow j \in \mathcal{P}. \end{cases} \quad (*)$$

Note that this will already hold for the DPAG $\tilde{\mathcal{P}}$ constructed by the first (arrowhead orientation) stage of FCI, i.e., after rules $\mathcal{R}0$–$\mathcal{R}4$ of the FCI algorithm (the only ones that can orient arrow heads) have been completed. It remains to show that if $k \rightarrow i$ in $\mathcal{P}$ for a third node k, then $k \rightarrow j$ in $\mathcal{P}$ as well (ruling out $k \circ\!\!\rightarrow j$). We will consider all FCI rules that can orient a tail at $k \rightarrow i$ in the absence of selection bias, i.e., FCI rules $\mathcal{R}1$, $\mathcal{R}4$a, $\mathcal{R}8$a, $\mathcal{R}9$, and $\mathcal{R}10$ in (Zhang, 2008b), and show that each of them implies also a tail at k on the edge to j, i.e., $k \rightarrow j$ (rules $\mathcal{R}5$–$\mathcal{R}7$ and $\mathcal{R}8$b can be ignored in the absence of selection bias). Below we use $\mathcal{P}'$ to denote an intermediate DPAG obtained so far by FCI during the orientation stage, which ultimately results in the completely oriented DPAG $\mathcal{P}$.

We will use the fact that there is a natural ordering in these orientation rules: $\mathcal{R}1$ and $\mathcal{R}4$a are part of the arrowhead orientation stage and complete first. Then, all instances of $\mathcal{R}9$ can be executed, after which $\mathcal{R}8$a and $\mathcal{R}10$ are triggered repeatedly until completion. The latter follows from the fact that rules $\mathcal{R}8$a (or $\mathcal{R}8$b) and $\mathcal{R}10$ can only orient an edge $x \circ\!\!\rightarrow y$ into $x \rightarrow y$, which can never introduce a new instance that satisfies the pattern of $\mathcal{R}9$ but did not already satisfy the pattern of $\mathcal{R}9$ before. We will assume (without loss of generality) that FCI makes use of this particular ordering in the proof below.

Rule $\mathcal{R}1$: *if $m \ast\!\!\rightarrow k \circ\!\!-\!\!\ast i$ in $\mathcal{P}'$ and m and i are not adjacent, then orient $m \ast\!\!\rightarrow k \rightarrow i$.* Suppose that $k \circ\!\!-\!\!\ast i$ in $\mathcal{P}'$ can been oriented by $\mathcal{R}1$. By $(*)$, this means that $k \ast\!\!\rightarrow j$ will be in $\tilde{\mathcal{P}}$. If $k \circ\!\!\rightarrow j$ would have remained unoriented in $\tilde{\mathcal{P}}$, then by Lemma 4 applied to $m \ast\!\!\rightarrow k \circ\!\!\rightarrow j$, there must be an edge $m \ast\!\!\rightarrow j$ in $\tilde{\mathcal{P}}$. But then there must also be an edge $m \ast\!\!\rightarrow i$ in $\tilde{\mathcal{P}}$, again by Lemma 4. This contradicts that m and i are not adjacent in $\mathcal{P}'$. Hence $k \rightarrow j$ must have been oriented in $\tilde{\mathcal{P}}$.

Rule $\mathcal{R}4$a: *if $\pi = \langle x, m_1, \ldots, m_n, k, i\rangle$ is a discriminating path for k in $\mathcal{P}'$ and $k \circ\!\!-\!\!\ast i$ is in $\mathcal{P}'$, and if $k \in \mathrm{SepSet}(x, i)$, then orient $k \rightarrow i$.* First note that j cannot be part of the discriminating path π, as $i \circ\!\!-\!\!\circ j$ in $\tilde{\mathcal{P}}$. By $(*)$, all nodes m_s and k also have an edge into j in $\tilde{\mathcal{P}}$ with either a tail or circle mark at the other end. So we have $x \ast\!\!\rightarrow m_1 \rightarrow j$ in $\tilde{\mathcal{P}}$ or $x \ast\!\!\rightarrow m_1 \circ\!\!\rightarrow j$ in $\tilde{\mathcal{P}}$. If x and j were adjacent in $\tilde{\mathcal{P}}$, then the edge between them must be of the form $x \ast\!\!\rightarrow j$ (either because of Lemma 4 if $m_1 \circ\!\!\rightarrow j$, or because of FCI rule $\mathcal{R}2$b if $m_1 \rightarrow j$), which would imply that also $x \ast\!\!\rightarrow i$ in $\tilde{\mathcal{P}}$ by $(*)$, contradicting the antecedent of rule $\mathcal{R}4$a. Now, by induction each edge between m_s and j (for $s = 1, \ldots, n$) will have been oriented as $m_s \rightarrow j$ in $\tilde{\mathcal{P}}$. Indeed, first $\mathcal{R}1$ can orient $x \ast\!\!\rightarrow m_1 \rightarrow j$, which means that $x \ast\!\!\rightarrow m_1 \leftarrow\!\!\ast\, m_2 \ast\!\!\rightarrow j$ is a discriminating path for m_2. Suppose $x \ast\!\!\rightarrow m_1(\leftrightarrow \ldots m_{s-1}) \leftarrow\!\!\ast\, m_s \ast\!\!-\!\!\ast j$ is a discriminating path for m_s with $s < n$. Now $m_s \leftrightarrow j$ cannot be in $\tilde{\mathcal{P}}$, because if it were, $m_s \leftrightarrow i$ would also be in $\tilde{\mathcal{P}}$ by $(*)$, contrary the antecedent of $\mathcal{R}4$a. Hence $m_s \in \mathrm{SepSet}(x, j)$, and so the edge between m_s and j can be oriented as $m_s \rightarrow j$ by $\mathcal{R}4$a, which means that $x \ast\!\!\rightarrow m_1(\leftrightarrow \ldots m_s) \leftarrow\!\!\ast\, m_{s+1} \ast\!\!-\!\!\ast j$ must be a discriminating path for m_{s+1}. Hence, $\pi' = \langle x, m_1, \ldots, m_n, k, j\rangle$ is also a discriminating path for k in $\tilde{\mathcal{P}}$, and again $k \leftrightarrow j$ cannot be in $\tilde{\mathcal{P}}$ (otherwise $k \leftrightarrow i$ in $\tilde{\mathcal{P}}$), and so the edge between k and j can be oriented by $\mathcal{R}4$a, resulting in $k \rightarrow j$ in $\tilde{\mathcal{P}}$.

Rule $\mathcal{R}9$: *if $k \circ\!\!\rightarrow i$ in $\mathcal{P}'$, and $\pi = \langle k, m_1, \ldots, m_n, i\rangle$ is an uncovered possibly directed path in $\mathcal{P}'$ from k to i such that m_1 and i are not adjacent in $\mathcal{P}'$, then orient $k \rightarrow i$.*

First note that all nodes on π must be ancestor in $\mathcal{G}$ of i, and i must be non-ancestor in $\mathcal{G}$ of all other nodes on π. Indeed, for any DMAG $\mathcal{H}$ contained in $\mathcal{P}'$, each node m_l must be a non-collider on π (all unshielded colliders have already been oriented by rule $\mathcal{R}0$ in $\mathcal{P}'$ by assumption, and there cannot be any on π since π is possibly directed). If $m_l \rightarrow m_{l+1}$ in $\mathcal{H}$, then there must be a directed path $m_l \rightarrow m_{l+1} \rightarrow \cdots \rightarrow m_n \rightarrow i$ in $\mathcal{H}$; if $m_l \rightarrow m_{l-1}$ in $\mathcal{H}$, then there must be a directed path $m_l \rightarrow m_{l-1} \rightarrow \cdots \rightarrow m_1 \rightarrow k \rightarrow i$ in $\mathcal{H}$. In both cases, $m_l \in \mathrm{AN}_{\mathcal{H}}(i)$, and $i \notin \mathrm{AN}_{\mathcal{H}}(m_l)$. Since this holds for any DMAG $\mathcal{H}$ contained in $\mathcal{P}'$, it also holds for all DMAGs induced by acyclifications of $\mathcal{G}$. Hence, by Proposition 3, $m_l \in \mathrm{AN}_{\mathcal{G}}(i)$, and $i \notin \mathrm{AN}_{\mathcal{G}}(m_l)$. This also implies (by the arrowhead completeness of FCI) that there must be an arrowhead on the edge $m_n \ast\!\!\rightarrow i$ in $\mathcal{P}'$.

By $(*)$, $k \ast\!\!\rightarrow j$ will be in $\mathcal{P}$. Both m_1 and m_{n-1} are not adjacent to j in $\mathcal{P}$. This follows from the fact that j cannot be ancestor of either of these nodes m_1, m_{n-1} in $\mathcal{G}$, for then i would also be ancestor in $\mathcal{G}$ of that node. Therefore an edge between m_1 (or m_{n-1}) and j in $\mathcal{P}$ would have to be into j by the arrowhead completeness of FCI, and so by $(*)$ there would also be an edge $m_1 \ast\!\!\rightarrow i$ (or $m_{n-1} \ast\!\!\rightarrow i$) in $\mathcal{P}$, contrary the antecedent of $\mathcal{R}9$. By $(*)$, also $m_n \ast\!\!\rightarrow j$ in $\mathcal{P}'$. This edge cannot be $m_n \leftrightarrow j$, because that would mean that also $m_i \leftrightarrow i$ in $\mathcal{P}'$ by $(*)$, a contradiction. Therefore $\pi' = \langle k, m_1, \ldots, m_n, j\rangle$

is also an uncovered possibly directed path from k to j in $\mathcal{P}'$, and m_1 and j are not adjacent in $\mathcal{P}'$, so $k \circ\!\!\rightarrow j$ can be oriented as $k \rightarrow j$ by $\mathcal{R}9$.

Finally, the two remaining rules, $\mathcal{R}8$a and $\mathcal{R}10$, will be considered together in order to be able to make use of a proof by induction. We will assume that in $\mathcal{P}'$, rules $\mathcal{R}1$, $\mathcal{R}4$a and $\mathcal{R}9$ have already been exhaustively applied.

Rule $\mathcal{R}8$a: *if $k \circ\!\!\rightarrow i$ and $k \rightarrow m \rightarrow i$ in $\mathcal{P}'$, then orient $k \rightarrow i$.*
Rule $\mathcal{R}10$: *if $k \circ\!\!\rightarrow i$ and $u_1 \rightarrow i \leftarrow u_2$ in $\mathcal{P}'$, $\pi_1 = \langle k, m_1, \ldots, u_1 \rangle$ is an uncovered possibly directed path from k to u_1 in $\mathcal{P}'$ and $\pi_2 = \langle k, m_2, \ldots, u_2 \rangle$ is an uncovered possibly directed path from k to u_2 in $\mathcal{P}'$, such that m_1 and m_2 are distinct and not adjacent, then orient $k \rightarrow i$.*

By $(*)$, if rule $\mathcal{R}8$a triggers in $\mathcal{P}'$, then also $k *\!\!\rightarrow j$ and $m *\!\!\rightarrow j$ in $\mathcal{P}$. Furthermore, if rule $\mathcal{R}10$ triggers in $\mathcal{P}'$, then also $k *\!\!\rightarrow j$, $u_1 *\!\!\rightarrow j$ and $u_2 *\!\!\rightarrow j$ in $\mathcal{P}$. As the arrowhead stage has already completed by the time $\mathcal{R}8$a and $\mathcal{R}10$ are executed, that means these will be present as edges into j in $\mathcal{P}'$ as well.

We now proceed by contradiction. Assume, for the sake of contradiction, that $\mathcal{R}8$a or $\mathcal{R}10$ triggers to orient some $k \circ\!\!\rightarrow i$ in $\mathcal{P}'$ as $k \rightarrow i$, but the corresponding edge $k \circ\!\!\rightarrow j$ (with $j \in \mathrm{SC}_{\mathcal{G}}(i)$) remains unoriented in $\mathcal{P}$. Consider the first edge $k \rightarrow i$ for which this situation occurs (in the sequence of orientations performed by FCI during this last part of the orientation phase).

If $k \circ\!\!\rightarrow i$ can be oriented by rule $\mathcal{R}8$a, then the edge $m \rightarrow i$ is already present in $\mathcal{P}'$ at that point. If the tail on that edge was oriented by one of the rules $\mathcal{R}1$, $\mathcal{R}4$a or $\mathcal{R}9$, also $m \rightarrow j$ will have been oriented in $\mathcal{P}'$ at this point, as we have shown. Otherwise, the tail of $m \rightarrow i$ must have been oriented by rule $\mathcal{R}8$a or rule $\mathcal{R}10$. By assumption, the corresponding edge $m \circ\!\!\rightarrow j$ does not remain unoriented in $\mathcal{P}$, and therefore after finitely many applications of rules $\mathcal{R}8$a and $\mathcal{R}10$, it will have been oriented as $m \rightarrow j$ in $\mathcal{P}'$. In both cases, the edge $m \rightarrow j$ will be present in $\mathcal{P}'$ at some point, and then rule $\mathcal{R}8$a can also be used to orient $k \circ\!\!\rightarrow j$ as $k \rightarrow j$, contradicting the assumption that $k \circ\!\!\rightarrow j$ remains unoriented in $\mathcal{P}$.

Hence $k \circ\!\!\rightarrow i$ must have been oriented by rule $\mathcal{R}10$. The edges $u_1 \rightarrow i$ and $u_2 \rightarrow i$ must then already be present in $\mathcal{P}'$ at that point. By similar reasoning as before, we conclude that also $u_1 \rightarrow i$ and $u_2 \rightarrow i$ must be present in $\mathcal{P}'$ at some point, and then rule $\mathcal{R}10$ can also be used to orient $k \circ\!\!\rightarrow j$ as $k \rightarrow j$, again contradicting the assumption.

Therefore, it cannot happen that $\mathcal{R}8$a or $\mathcal{R}10$ can be used to orient $k \circ\!\!\rightarrow i$ as $k \rightarrow i$, but that the corresponding edge $k \circ\!\!\rightarrow j$ (with $j \in \mathrm{SC}_{\mathcal{G}}(i)$) will not get oriented as $k \rightarrow j$ in $\mathcal{P}$.

Summarizing, whenever FCI orients a tail mark at k on an edge into i, it will also orient all tail marks at k on the edges into j for all $j \in \mathrm{SC}_{\mathcal{G}}(i)$. □

Proposition 10 *Let $\mathcal{G}$ be a DMG and denote by $\mathcal{P} = \mathcal{P}_{FCI}(\mathrm{IM}_\sigma(\mathcal{G}))$ the corresponding complete DPAG output by FCI. Let $i \neq j$ be two nodes in $\mathcal{P}$. If there is an edge $i \circ\!\!-\!\!\circ j$ in $\mathcal{P}$, and all nodes k for which $k *\!\!\rightarrow i$ is in $\mathcal{P}$ also have an edge of the same type $k *\!\!\rightarrow j$ (i.e., the two edge marks at k are the same) in $\mathcal{P}$, then there exists a DMG $\tilde{\mathcal{G}}$ with $j \in \mathrm{SC}_{\tilde{\mathcal{G}}}(i)$ that is σ-Markov equivalent to $\mathcal{G}$, but also a DMG $\mathcal{H}$ with $j \notin \mathrm{SC}_{\mathcal{H}}(i)$ that is σ-Markov equivalent to $\mathcal{G}$.*

Proof of Proposition 10 We first show that there is a DMAG $\mathcal{H}$ that is σ-Markov equivalent to $\mathcal{G}$ which has $i \rightarrow j$. We construct $\mathcal{H}$ by starting from the so-called arrowhead augmented graph (Zhang, 2006) of $\mathcal{P}$, in which each edge $x \circ\!\!\rightarrow y$ in $\mathcal{P}$ is oriented as $x \rightarrow y$, followed by an orientation of the remaining circle component into an arbitrary DAG such that no unshielded colliders are introduced. By Lemma 4.3.6 in (Zhang, 2006), this procedure yields a DMAG $\mathcal{H}$ in $\mathcal{P}$ (note that we assumed no selection bias), and hence σ-Markov equivalent to $\mathcal{G}$. Orienting the circle component (which is chordal by Lemma 4.3.7 in (Zhang, 2006)) can be achieved by choosing $i \rightarrow j$ as the top two root nodes in the partial ordering for the DAG, and propagating the Meek rules (Meek, 1995a) to orient the rest of the circle component into a DAG with no unshielded colliders (see also the discussion of "Meek's Algorithm" on pages 120–121 of (Zhang, 2006)). The resulting DMAG $\mathcal{H}$ has no other arrowheads at i and j other than the ones already present in $\mathcal{P}$, in combination with the directed edge $i \rightarrow j$.

We now create a DMG $\tilde{\mathcal{G}}$ out of $\mathcal{H}$ by adding an additional edge $j \rightarrow i$, thereby creating a single non-trivial strongly connected component containing only the nodes $\{i, j\}$. This follows from the fact that if there exists another node in $\mathrm{SC}_{\tilde{\mathcal{G}}}(i)$ as a result of adding the edge $j \rightarrow i$ to $\mathcal{H}$, then there must now be a directed path $j \rightarrow i \rightarrow \cdots \rightarrow k \rightarrow j$ in $\tilde{\mathcal{G}}$. However, this edge $k \rightarrow j$ cannot be part of the original circle component containing $i \circ\!\!-\!\!\circ j$ in $\mathcal{P}$, for then by construction it would have been oriented as $j \rightarrow k$, and if $k \rightarrow j$ was already an invariant arrowhead in $\mathcal{P}$ then by assumption there would also be an invariant arrowhead on $k *\!\!\rightarrow i$ in $\mathcal{P}$, which would imply the existence of an (almost) directed cycle $i \rightarrow \cdots \rightarrow k$ in combination with $k *\!\!\rightarrow i$ in the DMAG $\mathcal{H}$, which is impossible by definition.

As final step in the proof we show that the DMAG $\mathcal{H}$

qualifies as an acyclification of $\tilde{\mathcal{G}}$. By construction $\mathcal{H}$ is a DMAG, and also an ADMG, with the same vertices as $\tilde{\mathcal{G}}$. Since $j \in \mathrm{SC}_{\tilde{\mathcal{G}}}(i)$ and $i \to j$ is in $\mathcal{H}$, criterion (iii) of Definition 3 is satisfied. Finally, note that for all other nodes $k \notin \mathrm{SC}_{\tilde{\mathcal{G}}}(i)$ by assumption, if $k \to i$ in $\mathcal{P}$, then $k \to j$ is also in $\mathcal{P}$, and similarly for $k \leftrightarrow i$ and $k \leftrightarrow j$, resp. $k \circ\!\!\to i$ and $k \circ\!\!\to j$. By the initial step of the construction of $\mathcal{H}$ as the arrowhead augmented DMAG of $\mathcal{P}$, any edge $k \circ\!\!\to i$ in $\mathcal{P}$ is oriented as $k \to i$ in $\mathcal{H}$. As a result, for any node $k \notin \mathrm{SC}_{\tilde{\mathcal{G}}}(i)$, $k \to i$ in $\mathcal{H}$ iff $k \to j$ in $\mathcal{H}$, and similarly, $k \leftrightarrow i$ in $\mathcal{H}$ iff $k \leftrightarrow j$ in $\mathcal{H}$. Since $\{i, j\}$ is the only non-trivial strongly connected component in $\tilde{\mathcal{G}}$, criterion (ii) of Definition 3 is also satisfied. This proves that DMAG $\mathcal{H}$ indeed qualifies as an acyclification of $\tilde{\mathcal{G}}$.

This means that $\tilde{\mathcal{G}}$ is σ-Markov equivalent to $\mathcal{G}$, and hence, both $\tilde{\mathcal{G}}$ and $\mathcal{H}$ satisfy all properties promised in the claim of the proposition. $\square$

Theorem 2 *Let Ψ be background knowledge that satisfies Assumption 1 and let Φ be a mapping from independence models to DPAGs. Then:*

(i) If Φ is sound for background knowledge Ψ under the additional assumption of acyclicity, then Φ is sound for background knowledge Ψ.
(ii) If Φ is arrowhead (tail) complete for background knowledge Ψ under the additional assumption of acyclicity, then Φ is arrowhead (tail) complete for background knowledge Ψ.
(iii) If Φ is sound and arrowhead complete for background knowledge Ψ under the additional assumption of acyclicity, then Φ is Markov complete.

Proof of Theorem 2 For a DMG $\mathcal{G}$ with $\Psi(\mathcal{G}) = 1$, define

$$\mathrm{acy}(\mathcal{G}, \Psi) := \{\mathcal{G}' : \mathcal{G}' \text{ is an acyclification of } \mathcal{G} \text{ and } \Psi(\mathcal{G}') = 1\}.$$

(i) Suppose that Φ is sound for background knowledge Ψ under the additional assumption of acyclicity. Let $\mathcal{G}$ be a DMG with $\Psi(\mathcal{G}) = 1$ and let $\mathcal{P} := \Phi(\mathrm{IM}_\sigma(\mathcal{G}))$. For any $\mathcal{G}' \in \mathrm{acy}(\mathcal{G}, \Psi)$, which is nonempty by Assumption 1(i), $\mathcal{P} = \Phi(\mathrm{IM}_\sigma(\mathcal{G}'))$ contains $\mathcal{G}'$ by virtue of the acyclic soundness of Φ under background knowledge Ψ. Hence, two nodes u, v in $\mathcal{P}$ are adjacent if and only if there is an inducing path between u and v in $\mathcal{G}'$, and by Proposition 3(iii), this holds if and only if there is an inducing path between u and v in $\mathcal{G}$. Further, $u \ast\!\!\to v$ in $\mathcal{P}$ implies $v \notin \mathrm{AN}_{\mathcal{G}'}(u)$ for all $\mathcal{G}' \in \mathrm{acy}(\mathcal{G}, \Psi)$, and hence by Assumption 1(ii), this implies $v \notin \mathrm{AN}_{\mathcal{G}}(u)$. Finally, $u \to v$ in $\mathcal{P}$ implies $u \in \mathrm{AN}_{\mathcal{G}'}(v)$ for all $\mathcal{G}' \in \mathrm{acy}(\mathcal{G}, \Psi)$, and hence by Assumption 1(iii), this implies $u \in \mathrm{AN}_{\mathcal{G}}(v)$. We conclude that Φ is sound for background knowledge Ψ.

(ii) We give the proof for arrowhead completeness (tail completeness is proved similarly). Suppose that Φ is arrowhead complete for background knowledge Ψ under the additional assumption of acyclicity. I.e., for all ADMGs $\mathcal{G}'$ with $\Psi(\mathcal{G}') = 1$, any ancestral relation absent in any ADMG $\tilde{\mathcal{G}}'$ with $\Psi(\tilde{\mathcal{G}}') = 1$ that is d-Markov equivalent to $\mathcal{G}'$, is oriented as an arrowhead in $\Phi(\mathrm{IM}_d(\mathcal{G}'))$. Let $\mathcal{G}$ be a DMG with $\Psi(\mathcal{G}) = 1$ and let $\mathcal{P} := \Phi(\mathrm{IM}_\sigma(\mathcal{G}))$. Let $\mathcal{G}' \in \mathrm{acy}(\mathcal{G}, \Psi)$, which is nonempty by Assumption 1(i). Assume that an ancestral relation is absent in any DMG $\tilde{\mathcal{G}}$ with $\Psi(\tilde{\mathcal{G}}) = 1$ that is σ-Markov equivalent to $\mathcal{G}$. Then in particular, this ancestral relation is absent in any ADMG $\tilde{\mathcal{G}}'$ with $\Psi(\tilde{\mathcal{G}}') = 1$ that is d-Markov equivalent to $\mathcal{G}'$ (which is σ-Markov equivalent to $\mathcal{G}$). By the acyclic arrowhead completeness of Φ under background knowledge Ψ, it must be oriented as an arrowhead in $\Phi(\mathrm{IM}_d(\mathcal{G}')) = \mathcal{P}$. Hence Φ is arrowhead complete for background knowledge Ψ.

(iii) Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two DMGs that satisfy the background knowledge, i.e., $\Psi(\mathcal{G}_1) = \Psi(\mathcal{G}_2) = 1$. We have to show that if $\mathcal{G}_1$ and $\mathcal{G}_2$ are not σ-Markov equivalent, then $\Phi(\mathrm{IM}_\sigma(\mathcal{G}_1)) \neq \Phi(\mathrm{IM}_\sigma(\mathcal{G}_2))$. By Assumption 1(i), there exist acyclifications $\mathcal{G}_1' \in \mathrm{acy}(\mathcal{G}_1, \Psi)$ and $\mathcal{G}_2' \in \mathrm{acy}(\mathcal{G}_2, \Psi)$. Proposition 2 implies both $\mathrm{IM}_\sigma(\mathcal{G}_1) = \mathrm{IM}_d(\mathcal{G}_1')$ and $\mathrm{IM}_\sigma(\mathcal{G}_2) = \mathrm{IM}_d(\mathcal{G}_2')$. Assume that $\mathcal{G}_1$ and $\mathcal{G}_2$ are not σ-Markov equivalent. Then their acyclifications $\mathcal{G}_1'$ and $\mathcal{G}_2'$ are not d-Markov equivalent. By assumption, both acyclifications $\mathcal{G}_1', \mathcal{G}_2'$ satisfy the background knowledge, i.e., $\Psi(\mathcal{G}_1') = \Psi(\mathcal{G}_2') = 1$.

The induced DMAGs $\mathcal{H}_1 := \mathrm{DMAG}(\mathcal{G}_1')$ and $\mathcal{H}_2 := \mathrm{DMAG}(\mathcal{G}_2')$ cannot be d-Markov equivalent because the induced DMAGs preserve the conditional independence models. Therefore, by the result of Ali et al. (2009), $\mathcal{H}_1$ and $\mathcal{H}_2$ either have a different skeleton, or they have the same skeleton but different colliders with order. If their skeletons differ, then also $\Phi(\mathrm{IM}_d(\mathcal{G}_1')) \neq \Phi(\mathrm{IM}_d(\mathcal{G}_2'))$ by the assumed soundness of Φ. If they do have the same skeletons, there must be at least one collider with order in $\mathcal{H}_1$ that is not a collider with order in $\mathcal{H}_2$, or vice versa. By Lemma 3.13 in Ali et al. (2009), this would imply that there is at least one collider with order in $\mathcal{H}_1$ that is not a collider in $\mathcal{H}_2$, or vice versa.

Without loss of generality, assume that the former holds. The assumed soundness and arrowhead completeness of Φ under the additional assumption of acyclicity imply that this collider with order in $\mathcal{H}_1 = \mathrm{DMAG}(\mathcal{G}_1')$ will appear as a collider in $\Phi(\mathrm{IM}_d(\mathcal{G}_1'))$. Also, the soundness of Φ under the additional assumption of acyclicity implies that this noncollider in $\mathcal{H}_2$ cannot end up as a collider in $\Phi(\mathrm{IM}_d(\mathcal{G}_2'))$.

Hence, $\Phi(\mathrm{IM}_d(\mathcal{G}'_1)) \neq \Phi(\mathrm{IM}_d(\mathcal{G}'_2))$, and therefore, $\Phi(\mathrm{IM}_\sigma(\mathcal{G}_1)) \neq \Phi(\mathrm{IM}_\sigma(\mathcal{G}_2))$. $\square$

Lemma 3 *If subset $\{1\}$, $\{1,2\}$, or $\{1,2,3\}$ of the JCI Assumptions holds for a DMG $\mathcal{G}$, then the same subset of assumptions holds for any acyclification of $\mathcal{G}$.*

Proof of Lemma 3 Let $\mathcal{G}'$ be an acyclification of $\mathcal{G}$. JCI Assumption 1 implies that each strongly connected component in $\mathcal{G}$ consists entirely of system variables or entirely of context variables. Since in addition, $\mathcal{G}$ does not have any directed edge from a system to a context variable, there will not be any spurious directed edge in $\mathcal{G}'$ from a system to a context variable. Hence also $\mathcal{G}'$ satisfies JCI Assumption 1. If $\mathcal{G}$ satisfies also JCI Assumption 2, the acyclification $\mathcal{G}'$ will not contain any spurious bidirected edge between a context and a system variable. Hence $\mathcal{G}'$ satisfies JCI Assumptions 1 and 2 if $\mathcal{G}$ does so. Finally, it is clear that JCI Assumption 3 holds for $\mathcal{G}'$ if JCI Assumptions 1 and 3 hold for $\mathcal{G}$. $\square$

Corollary 4 *Let $\mathcal{M}$ be a simple (possibly cyclic) SCM with graph $\mathcal{G}(\mathcal{M})$ and assume that its distribution $\mathbb{P}_\mathcal{M}(\boldsymbol{X})$ is σ-faithful w.r.t. the graph $\mathcal{G}(\mathcal{M})$. When using consistent conditional independence tests on an i.i.d. sample of observational data from the induced distribution $\mathbb{P}_\mathcal{M}(\boldsymbol{X})$ of $\mathcal{M}$, FCI provides a consistent estimate $\hat{\mathcal{P}}$ of the DPAG $\mathcal{P}_{\scriptscriptstyle FCI}(\mathrm{IM}_\sigma(\mathcal{G}(\mathcal{M})))$ that represents the σ-Markov equivalence class of $\mathcal{G}(\mathcal{M})$. From the estimated DPAG $\hat{\mathcal{P}}$, we obtain consistent estimates for: (i) the absence/presence of (possibly indirect) causal relations according to $\mathcal{M}$ via Propositions 4 and 5; (ii) the absence of confounders according to $\mathcal{M}$ via Proposition 6; (iii) the absence/presence of direct causal relations according to $\mathcal{M}$ via Propositions 7 and 8; (iv) the absence of causal cycles according to $\mathcal{M}$ via Proposition 9.*

Proof of Corollary 4 In general, soundness of a constraint-based causal discovery algorithm (i.e., the correctness of its output when given the true independence model as input) implies consistency of the algorithm when using appropriate conditional independence tests. $\square$

Proposition 12 *Let $\mathcal{G}$ be a DMG that satisfies JCI Assumptions $\{1,2,3\}$. Let $\mathcal{P} = \mathcal{P}_{\scriptscriptstyle FCI_{JCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$ denote the DPAG output by the corresponding version of FCI-JCI. Let $i \in \mathcal{K}$, $j \in \mathcal{I}$. Then:*

1. *If i is not adjacent to j in $\mathcal{P}$, $i \to j$ is not in $\mathcal{G}$.*
2. *If $i \to j$ in $\mathcal{P}$, and for all system nodes $k \in \mathcal{I}$ s.t. $i \to k$ in $\mathcal{P}$ and k ∘–∘ j or k ∘→ j or $k \to j$ in $\mathcal{P}$, the edge $k \to j$ is definitely visible in the DPAG $\mathcal{P}^*$ obtained from $\mathcal{P}$ by replacing the edge between k and j by $k \to j$, then $i \to j$ is present in $\mathcal{G}$.*

Proof of Proposition 12 By Corollary 2, $\mathcal{P}$ contains $\mathcal{G}$. The first statement then follows directly from Proposition 7.

For the second statement, note first that there must be an inducing walk π between i and j in $\mathcal{G}$ that is into j by Lemma 1. We will show that any such inducing walk cannot be into i. On the contrary, suppose that π would be into i. Because of JCI Assumptions 1 and 2, the node on π next to i cannot be in $\mathcal{I}$ but must be a context node in $\mathcal{K}$. Similarly, all subsequent nodes on the inducing path (except for the final node j) must be collider nodes in $\mathcal{K}$ because of JCI Assumptions 1, 2 and 3. But then the final edge of π is between a context node and system node i and into the context node, contradicting JCI Assumption 1 or 2.

The claim now follows from the second statement of Proposition 8. $\square$

Proposition 13 *Let $\mathcal{G}$ be a DMG that satisfies JCI Assumptions $\{1,2,3\}$. Let $\mathcal{P} = \mathcal{P}_{\scriptscriptstyle FCI_{JCI}}(\mathrm{IM}_\sigma(\mathcal{G}))$ denote the complete DPAG output by the corresponding version of FCI-JCI. Let $i \neq j$ be two nodes in $\mathcal{P}$. If $j \in \mathrm{sc}_\mathcal{G}(i)$, then i ∘–∘ j in $\mathcal{P}$, and for all nodes $k \neq i,j$: $k \to i$ in $\mathcal{P}$ iff $k \to j$ in $\mathcal{P}$, and $k \leftrightarrow i$ in $\mathcal{P}$ iff $k \leftrightarrow j$ in $\mathcal{P}$, and k ∘→ i in $\mathcal{P}$ iff k ∘→ j in $\mathcal{P}$.*

Proof of Proposition 13 We make use of a similar strategy as in the proof of (Theorem 38 in Mooij et al., 2020), which we will not repeat here. Proposition 9 applies to the extended complete DPAG $\mathcal{P}^*$ constructed in that proof. Since strongly connected components can only occur amongst the system variables, the same orientations will be found in the DPAG $\mathcal{P}$ output by FCI-JCI. $\square$