Structure Learning for Unfaithful Distributions: The Minimal Dependence Faithfulness

Pouria Ramazi[1] **Hamid Kalantari**[2]

[1]Departments of Mathematics and Statistics and Biological Sciences, University of Calgary, Calgary, Alberta, Canada
[2]Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

Abstract

Causal discovery algorithms such as PC often rely on faithfulness; yet XOR-type relations expose violations that PC and its weakened-faithfulness variants can miss. In such cases, a target variable X may be independent of each variable Y_i individually, while becoming dependent on Y_i once the remaining variables are conditioned on. We call a minimal set with this property a minimal dependence (MD) set: no proper subset has the same dependence pattern. MD sets of size at least two violate faithfulness. We introduce minimal-dependence faithfulness and a corresponding orientation condition, characterize the resulting DAG structure by showing that dependent members of an MD set connect to X directly or through colliders, and define the associated MD-equivalence representation. We then present MD-PC, a PC-style algorithm that detects such faithfulness violations and, under the proposed assumptions, returns the corresponding candidate class of DAGs.

1 INTRODUCTION

Structural causal models (SCMs) are used to capture the causal relationships between a set of random variables. They assign to each variable a deterministic function of some of the other variables, known as its *parents* or *direct causes*, and possibly a noise variable [Peters et al., 2017]. Consequently, each SCM induces a joint probability distribution over the random variables and a directed acyclic graph (DAG) where the direct causes of each variable are linked to that variable. The DAG is also known as the "true DAG" or *causal network* of the variables. To learn the structure of the causal network – a process referred to as *structure learning* – intervention is a reliable method to find causal relations [Ke et al., 2020]. However, because of practical constraints, the use of only observational data has attracted attention in recent decades [Ng et al., 2021] and Roohi et al. [2021]. For this purpose, sufficiency, causal Markov condition (CMC), and faithfulness are usually considered in the literature to make learning possible [Zhang and Spirtes, 2016]. Causal sufficiency refers to the absence of hidden (or latent) variables Sanjaroonpouri and Ramazi [2025]. CMC implies that all conditional independencies implied by the true DAG hold in the joint probability distribution and faithfulness implies the inverse.

Under these assumptions, one main approach to find the true DAG is known as *constraint-based* and is based on testing conditional independencies (constraints) between the variables [Kitson et al., 2021]. Examples of constraint-based algorithms are PC, SGS, and FCI [Spirtes et al., 2000] or their scalable versions Kalantari et al. [2025] which are based on (in)dependence detection. The result of this approach is a class of independence-equivalent (I-equivalent) graphs that are presented as a partially DAG (PDAG). Under the causal Markov and faithfulness assumptions, constraint-based approaches have been shown to correctly find the true PDAG given the joint probability distribution, and the same holds asymptotically with respect to the number of data instances of a given dataset [Ng et al., 2021]. CMC and faithfulness are used to conclude the existence and absence of links in the true DAG based on conditional independence test results, respectively. However, the faithfulness assumption is violated in some practical situations [Andersen, 2013], making it a restrictive assumption. Some of the work in the literature has accordingly relaxed this assumption [Lemeire et al., 2012] and developed *unfaithfulness detection* approaches, that is, to detect from observational distribution P the existence of some conditional dependencies that cannot be captured by the true DAG [Raskutti and Uhler, 2018b, Kocaoglu, 2023, Wienöbst and Liskiewicz, 2020, Sadeghi, 2017].

Adjacency-faithfulness and orientation-faithfulness assumptions that are weaker than faithfulness were defined in [Ramsey et al., 2006]. PC algorithm was accordingly adapted, resulting in the *conservative PC (CPC) algorithm* that puts

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:5661–5676.

forward a number of candidate PDAGs that include the true DAG. As another relaxation of faithfulness, triangle-faithfulness was introduced and investigated in [Zhang and Spirtes, 2008, Spirtes and Zhang, 2014], resulting in the *very conservative PC (VCPC)* and *VCSGS* algorithms. However, all of these and other existing algorithms, such as MGM-FCI-MAX [Horii, 2021], fail to detect the true graph in examples such as the generalized XOR where several Bernoulli random variables cause another variable. The first step in detecting this type of unfaithfulness was taken in [Marx et al., 2021] for the case with three variables, i.e., two causing the third, resulting in the conventional XOR example. The authors relaxed the faithfulness to the so-called 2-adjacency faithfulness and 2-orientation faithfulness, provided a modified grow-shrink (GS) algorithm [Margaritis and Thrun, 1999a] to detect triples of variables violating faithfulness, and proved the soundness.

We take the next step in this regard and define the *minimal dependence* of a given variable X as the set of variables Y, such that X is independent of each Y but becomes dependent if conditioned on the remainder of the set. Minimal dependencies of size at least two violate faithfulness. Consequently, we relax the assumption to *minimal dependence faithfulness*, limiting the existence of links to a variable and its minimal dependence. We then determine the structure of minimal dependencies in the true DAG and provide a modification of the PC algorithm, the minimal dependence PC (MD-PC) algorithm to detect this kind of unfaithfulness and output all possible candidates for the true DAG.

2 PROBLEM FORMULATION

Consider a *structural causal model (SCM)* [Peters et al., 2017] defined as the tuple $\langle \mathcal{X}, \mathcal{N}, P_{\mathcal{U}}, \mathcal{F} \rangle$ where $\mathcal{X} = \{X_1, \ldots, X_m\}$ are the random variables, $\mathcal{N} = \{N_X \mid X \in \mathcal{X}\}$ are the disjoint noise random variables whose joint distribution $P_{\mathcal{U}}$ satisfies $P_{\mathcal{U}}(\mathcal{N}) = \prod_X P_{\mathcal{U}}(N_X)$. The value assigned to each variable $X \in \mathcal{X}$ is a deterministic function f_X of a subset of the variables $\mathrm{Pa}_X \subset \mathcal{X} \setminus \{X\}$, known as its *parents*, and a parent noise variable $N_X \in \mathcal{N}$, i.e., $X := f_X(\mathrm{Pa}_X, N_X)$. The set of functions f_X for all variables X defines $\mathcal{F}$. The parents of each variable X are known as its *direct causes*. The function f_X is *minimal* in the parents of X; that is, there does not exist another function g_X, such that $f_X(\mathrm{Pa}_X, N_X) = g_X(\mathcal{S}, N_X)$ for some subset $\mathcal{S} \subset \mathrm{Pa}_X$. The SCM induces *(i)* a joint (observational) probability distribution P over the variables $\mathcal{X}$ and *(ii)* a DAG $\mathcal{G}$ whose nodes represent variables $\mathcal{X}$, and there is a link to every variable X from each of its parents Pa_X.

The goal is to obtain DAG $\mathcal{G}$, referred to as the "true DAG," from observational probability distribution P, that is to estimate the causal relationships from the observational distribution.

Problem 1 (Causal discovery) *Consider an SCM over the variables $\mathcal{X}$, inducing DAG $\mathcal{G}$ and joint observational probability distribution P. Given distribution P, find DAG $\mathcal{G}$.*

This is, however, an impossible task in general. The simplest example is the case with two dependent variables X and Y, where both SCMs $X := 2Y$ and $Y := 0.5X$ impose the same observational distribution P but two different DAGs $Y \to X$ and $X \to Y$.

Some commonly-imposed assumptions facilitate this task. One is the *(causal) faithfulness* assumption explained as follows. True DAG $\mathcal{G}$ implies some independencies between the variables. For example, if there is no trail (path) between two variables in $\mathcal{G}$, their deterministic functions are not related, rendering the two variables independent. Faithfulness states that such independencies implied by $\mathcal{G}$ capture all independencies satisfied by the distribution P. The attribution of (conditional) independencies to a DAG is formalized by the notion of *d-separation* (see Appendix). Let $\mathcal{I}(\mathcal{G})$ denote the set of all d-separations in a DAG $\mathcal{G}$ and $\mathcal{I}(P)$ denote the set of all conditional independencies implied by distribution P, i.e., $\mathcal{I}(P) = \{(\mathcal{Y}_1 \perp \mathcal{Y}_2 \mid \mathcal{Y}_3) : \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3 \subseteq \mathcal{X}\}$.

Assumption 1 (Faithfulness) $\mathcal{I}(P) \subseteq \mathcal{I}(\mathcal{G})$.

Distribution P is said to be *(causally) faithful* to the DAG $\mathcal{G}$ if it satisfies the above assumption. Problem 1 can then be stated as follows.

Problem 2 (Causal discovery under faithfulness) *Consider an SCM over variables $\mathcal{X}$, inducing DAG $\mathcal{G}$ and joint observational probability distribution P. Given distribution P and under Assumption 1, find DAG $\mathcal{G}$.*

Problem 2 is often introduced from a different perspective, without the causality ingredient: there is some true DAG $\mathcal{G}$ over random variables $\mathcal{X}$ with joint distribution P, and the goal is to again obtain $\mathcal{G}$ from P. Clearly, then the connection between $\mathcal{G}$ and P is lost and any DAG is a candidate for $\mathcal{G}$. Faithfulness helps to bridge the gap yet is insufficient as, for example, an empty graph satisfies the faithfulness assumption for every distribution P, and hence, is always a valid candidate. A second commonly-imposed assumption is the *(Causal) Markov Condition (CMC)*.

Assumption 2 (Markovness) $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(P)$.

If Markovness holds, $\mathcal{G}$ is called an *I-map* for P. If additionally, $\mathcal{G}$ is no longer an I-map for P upon the elimination of an edge, then it is a *minimal I-map* for P. Similar to faithfulness, Markovness is insufficient to learn causal DAGs from observational data, because, for example, a fully connected DAG has $\mathcal{I}(\mathcal{G}) = \emptyset$ and satisfies CMC for any observational distribution. However, together these assumptions enable us to present Problem 2 without the causality ingredient.

Problem 3 (Structure learning) *Consider random variables $\mathcal{X}$ with joint probability distribution P. Assume there exists a DAG $\mathcal{G}$ satisfying Assumptions 1 and 2. Given distribution P, find DAG $\mathcal{G}$.*

We state the results in the conventional setup of Problem 3, but they can also be stated using the setup of Problem 1.Note that even with both Markovness and faithfulness in force, the two-node DAGs example stated at the beginning of this section may not be distinguished using the observational distribution. This is because the two have identical distributions P and in turn $\mathcal{I}(P)$. Consequently, the problem of finding the true DAG is reduced to finding the so-called *I-equivalent* class of DAGs, which share the same $\mathcal{I}(G)$ that is equal to $\mathcal{I}(P)$. I-equivalent DAGs are known to have the same skeleton and same directions on some of the edges [Koller and Friedman, 2009], and hence, are represented by a PDAG.

PC algorithm solves Problem 3 (and 2) by putting forward an I-equivalence class that includes the true DAG $\mathcal{G}$. The algorithm, however, fails if the faithfulness assumption is violated. The idea with PC is that if two variables are independent conditioned on any subset of the other variables, then the two are not adjacent in the true DAG–a condition satisfied by faithfulness. If distribution P is unfaithful, PC may incorrectly identify the absence of a link once an independence is detected in P as illustrated in the following example.

Example 1 (n^{th}-order XOR) *Consider the SCM with variables $Y_1, \ldots, Y_n$, and X defined by*

$$Y_i := \text{Bernoulli}(1/2), \quad i = 1, \ldots, n,$$
$$X := Y_1 \oplus \ldots \oplus Y_n,$$

where $\oplus$ is the XOR *function. The induced causal DAG is shown in Figure 1(a). By probabilistic arithmetic, $P(X \mid Y_i) = P(X)$ for all i. Therefore, X is independent from every Y_i, but X is not independent of any Y_i conditioned on the remainder of Y_i's, i.e., $P(X \mid Y_1, \ldots, Y_n) \neq P(X \mid Y_i)$ for all i, implying the existence of some path between X and $Y_1, \ldots, Y_n$. Therefore, the independencies $X \perp Y_i$ for all i are satisfied by the distribution P but not by the causal DAG $\mathcal{G}$. As a result, faithfulness is violated, i.e., $\mathcal{I}(P) \nsubseteq \mathcal{I}(\mathcal{G})$.*

Should existing constraint-based algorithms, i.e., SGS and PC or their modified versions such as CSGS, CPC, VCSGS, or VCPC, be applied to the above observational distribution, the result would be the empty DAG $\mathcal{G}'$ in Figure 1(b), which is different from $\mathcal{G}$ and violates the causal Markov condition, i.e., $(X \perp Y_1 \mid Y_2, \ldots, Y_n) \in \mathcal{I}(\mathcal{G}')$ whereas $(X \perp Y_1 \mid Y_2, \ldots, Y_n) \notin \mathcal{I}(P)$.

To the best of our knowledge, except for the case of $n = 2$ [Marx et al., 2021], this type of unfaithfulness has not been investigated in detail and no constraint-based algorithm is able to detect it. The unfaithfulness is also not limited to Bernoulli distributions, binary variables, or deterministic relationships; see Examples 4 and 5 in the Appendix. It is also not limited to extended v-structures as in Figure 1(a) as illustrated in the following example. Given set of variables $\mathcal{Y} = \{Y_1, \ldots, Y_n\}, n \geq 1$, define $\mathcal{Y}_{-i} = \mathcal{Y} \setminus \{Y_i\}$ for $i = 1, \ldots, n$.

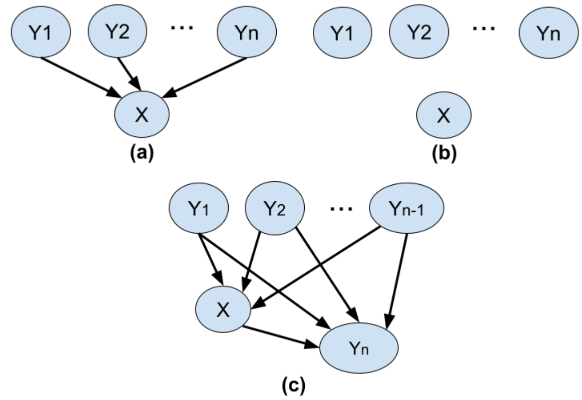

Figure 1: (a) The true DAG in Example 1, (b) the DAG found by PC and its modified versions, (c) the true DAG in Example 2

Example 2 *Consider the SCM with variables $Y_1, \ldots, Y_n$, and X defined by*

$$Y_i := \text{Bernoulli}(1/2), \quad i = 1, \ldots, n-1,$$
$$X := Y_1 \oplus \ldots \oplus Y_{n-1} \oplus U,$$
$$Y_n := Y_1 \oplus \ldots \oplus Y_{n-1} \oplus X,$$

where $U := \text{Bernoulli}(1/2)$ is the noise variable. The induced causal DAG $\mathcal{G}$ is shown in Figure 1(c). For every index i, it can be verified that $X \perp Y_i$, but $X \not\perp Y_i \mid \mathcal{Y}_{-i}$. Thus, similar to Example 1, the induced joint probability distribution P is unfaithful.

In what follows, we first characterize the type of unfaithfulness appearing in these examples, then determine the structure of the underlying true DAG, and finally propose an algorithm that identifies a class of DAGs that includes the true DAG.

3 MINIMAL-DEPENDENCE FAITHFULNESS

The type of unfaithfulness explained in Examples 1 and 2 happens when a variable X is independent of each individual member of a set $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ but not when conditioned on the other members. Namely, X indeed depends on all of $Y_1, \ldots, Y_n$, yet each single Y_i does not individually inform the distribution of X. The following definition captures this idea.

Definition 1 (Weak minimal dependence) *Given a variable $X \in \mathcal{X}$, a* weak minimal dependence *for X is a set $\mathcal{Y} \subseteq \mathcal{X} \setminus \{X\}$ that satisfies the following conditions:*

1. *there exists at least one member $Y \in \mathcal{Y}$ that satisfies*
$$X \not\perp Y \mid \mathcal{Y} \setminus \{Y\}, \tag{1}$$
2. *if $|\mathcal{Y}| \geq 2$, then*
$$\forall Y \in \mathcal{Y} \quad X \perp Y, \tag{2}$$
3. *no proper subset of $\mathcal{Y}$ satisfies* (1) *and* (2)*.*

If a minimal dependence of a variable X consists of a single variable Y, then Condition (2) does not apply. Condition (1), on the other hand, implies that X and Y are dependent. Should faithfulness be in force, we would conclude that X and Y are connected by some active trail.

However, if a minimal dependence entails at least two variables, it violates the faithfulness assumption. Condition (1) implies that X depends on some variable Y in its minimal dependence set given the remainder of the set. However, Condition (2) implies that X is independent of every member of the minimal dependence. This is counterintuitive as it implies that the only way the dependence between X and Y can be explained is by the other members of the minimal dependence; nevertheless, X is independent of each of these members, resulting in an unfaithful distribution.

Proposition 1 *Consider random variables $\mathcal{X}$ with joint probability distribution P. If for any of the variables $X \in \mathcal{X}$, there exists a minimal dependence of X with cardinality of at least two, then P is not faithful to any DAG $\mathcal{G}$ over $\mathcal{X}$ that is an I-map for P.*

A closely related and well-established setting arises in genetic interaction studies. In pure epistasis, or epistatic models without marginal effects, a phenotype may depend on a combination of genetic loci even though no individual locus has a marginal association with the phenotype. Such interaction-only structures can be missed by marginal screening. Purely epistatic models with no main effects have been formally characterized in the genetics literature Culverhouse et al. [2002], and multifactor dimensionality reduction was developed to detect high-order gene–gene and gene–environment interactions that may not be revealed by single-variable analyses Ritchie et al. [2001], Motsinger and Ritchie [2006]. Recent simulation work also explicitly considers epistatic models without marginal effects Shang et al. [2022]. Hence, these examples demonstrate that joint-only, or approximately joint-only, dependence is a practically recognized phenomenon. As a nonmedical illustration, Appendix Example 6 gives a water-tank process-control setting with the same qualitative behavior.

Example 3 (influenza or dengue fever) *Let the binary variable X indicate the presence of either influenza ($X = 0$) or the dengue fever disease [Halstead, 2007, Eccles, 2005] ($X = 1$) in a patient. There are three symptoms shared between these two diseases:* (i) *fever (captured by $Y_1 = 1$ when present otherwise $Y_1 = 0$);* (ii) *headache (captured by $Y_2 = 1$ when present otherwise $Y_2 = 0$);* (iii) *joint and muscle pain (captured by $Y_3 = 1$ when present otherwise $Y_3 = 0$). Knowing any of these symptoms does not make either of the two diseases more likely, i.e., $P(X \mid Y_i) = P(X)$ for all i. However, while all three symptoms can be simultaneously present in the case of influenza, that is not the case with dengue fever. In dengue fever, fever and headache appear in the early stages, and the joint and muscle pain is observed often a few days after. Therefore, if the patient is known to have fever and headache, the additional information of whether they also have joint and muscle pain may determine the disease type, i.e., $P(X \mid Y_1, Y_2, Y_3) \neq P(X \mid Y_1, Y_2)$, implying $X \not\perp Y_3 \mid Y_1, Y_2$. It follows that $\mathcal{Y} = \{Y_1, Y_2, Y_3\}$ is a minimal dependence for X.*

The set $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ in Examples 1 and 2 is a dependence set for X. Example 3 is intended as an intuitive illustration rather than as the sole practical motivation for minimal dependence.

The third condition in Definition 1 is to ensure that the members of $\mathcal{Y}$ are all required for the unfaithfulness; otherwise, the dependence set $\mathcal{Y}$ may become redundantly large by including variables that are independent of every other variable, e.g., isolated nodes. Nevertheless, the consequence of minimality is that all members of the minimal dependence set satisfy Condition 1.

Proposition 2 *Let $\mathcal{Y} \subseteq \mathcal{X}$ be a weak minimal dependence of a variable $X \in \mathcal{X}$. Then*
$$\forall Y \in \mathcal{Y} \quad X \not\perp Y \mid \mathcal{Y} \setminus \{Y\} \tag{3}$$
$$X \not\perp \mathcal{Y} \tag{4}$$
and for every $\mathcal{Y}' \subset \mathcal{Y}$,
$$X \perp \mathcal{Y}', \tag{5}$$
$$\forall Y \in \mathcal{Y}' \quad X \perp Y \mid \mathcal{Y}' \setminus \{Y\}. \tag{6}$$

Although we defined minimal dependence in Definition 1 in its simplest form by requiring the dependence in Condition (1) to hold for only one member, it appears that the condition does hold for all members of the minimal dependence. This means that in practice, if for a set of variables $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ that are each independent of X, we find that just one of them, say Y_1, depends on X conditioned on the rest of the variables, then so do all of the remainder, i.e., $Y_2, \ldots, Y_n$, provided that $\mathcal{Y}$ is minimal. If $\mathcal{Y}$ is not minimal, then we can find the minimal set by finding the smallest subset of $\mathcal{Y}$ that satisfies Condition (1). This explains why a uniform quantifier is not used in Condition (1).

3.1 THE STRUCTURE OF THE MINIMAL DEPENDENCE

What can be concluded about the true DAG $\mathcal{G}$ of a distribution P that admits a minimal dependence of size at least two? An immediate result of Proposition 1 is that once a minimal dependence of size at least two is detected for a variable X in the distribution P, the absence of the links between X and its minimal dependence variables may no longer be concluded, even though X is independent of them. This, however, does not mean that X and the minimal dependence can form any DAG. The following result enforces the existence of certain combinations of links.

Theorem 1 *Consider random variables $\mathcal{X}$ with joint probability distribution P. Consider a node $Z \in \mathcal{X}$ and its weak minimal dependence $\mathcal{Y}$. If DAG $\mathcal{G}$ is an I-map for P, then every member $Y \in \mathcal{Y}$ is connected to Z either by a collider-free trail or by a trail, every collider of which has a node in $\mathcal{Y}$ as its descendant.*

To further conclude about the structure of the true DAG, we define the following stronger version of Definition 1. This definition is motivated by normal faithfulness, where two nodes X and Y are adjacent if and only if their dependence does not break upon the observation of any other set $\mathcal{U}$, i.e., $X \not\perp Y \mid \mathcal{U}$ for all $\mathcal{U} \not\supseteq \{X, Y\}$ [Koller and Friedman, 2009]. Similarly, here, we extend Condition (1) in Definition 1 to when any additional set $\mathcal{U}$ is observed.

Definition 2 ((Strong) minimal dependence) *Given a variable $X \in \mathcal{X}$, a* (strong) minimal dependence *for X is a set $\mathcal{Y} \subseteq \mathcal{X} \setminus \{X\}$ that satisfies the following conditions:*

1. *there exists at least one member $Y \in \mathcal{Y}$ that satisfies*

$$\forall \mathcal{U} \subset \mathcal{X} \backslash (\{X\} \cup \mathcal{Y}) \qquad X \not\perp Y \mid (\mathcal{Y} \backslash \{Y\}) \cup \mathcal{U} \quad (7)$$

2. *if $|\mathcal{Y}| \geq 2$, then*

$$\forall Y \in \mathcal{Y} \quad X \perp Y, \qquad (8)$$

3. *no proper subset of $\mathcal{Y}$ satisfies* (7) *and* (8).

Every member $Y \in \mathcal{Y}$ that satisfies (7) *is called a* dependent member *of $\mathcal{Y}$ (for X) and the set of all dependent members of $\mathcal{Y}$ is denoted by $\mathbf{Dep}^{\circ}(\mathcal{Y}, X)$ (and when clear from the context, $\mathcal{Y}^{\circ}$). The set of all strong minimal dependencies of X is denoted by $\mathbf{Dep}(X)$.*

In the rest of the paper, by "minimal dependence" we mean strong minimal dependence. According to Definition 2, the dependence between X and its minimal dependence variables does not break upon the observation of non-minimal-dependence variables. Clearly, a strong minimal dependence of a variable X is also its weak minimal dependence but not the other way around. Note that (7) implies the presence of an edge between X and a node in $\mathcal{Y}$.

Next, to deduce the absence of an edge in the true DAG, we make the following assumption. The violation of faithfulness no longer allows edge elimination in $\mathcal{G}$. Nevertheless, faithfulness is only violated "within" the minimal dependence sets. The outer ones may be removed.

Assumption 3 (Minimal-dependence faithfulness) *Consider DAG $\mathcal{G}$ over variables $\mathcal{X}$ with joint probability distribution P. If $Y, Z \in \mathcal{X}$ are adjacent in $\mathcal{G}$, then each is a dependent member of a minimal dependence of the other, i.e., there exists a minimal dependence $\mathcal{Y} \in \mathbf{Dep}(Z)$ such that $Y \in \mathcal{Y}^{\circ}$ and a minimal dependence $\mathcal{Z} \in \mathbf{Dep}(Y)$ such that $Z \in \mathcal{Z}^{\circ}$.*

This assumption is motivated by a result in faithfulness (as well as adjacency faithfulness [Ramsey et al., 2006]; see Section 3.4). Under a faithful distribution, two nodes Y and Z are adjacent only if $Y \not\perp Z \mid \mathcal{U}$ for all $\mathcal{U} \subseteq \mathcal{X} \setminus \{Y, Z\}$. Namely, adjacency is allowed only between "tightly" dependent variables. Similarly, Assumption 3 enforces the same, with the difference that here the notion of tight dependence extends to that between a variable Z and a **set** of other variables $\mathcal{Y}$. Namely, in view of Proposition 2, each minimal dependence $\mathcal{Y} \in \mathbf{Dep}(Z)$ can be seen as a "super node": Although independent of individual nodes in $\mathcal{Y}$, variable Z does depend on the whole $\mathcal{Y}$ according to Equation (7). Hence, Assumption 3 allows adjacency between nodes Z and Y only if Z and $\mathcal{Y}$ are tightly dependent where $Y \in \mathcal{Y}^{\circ}$ and $\mathcal{Y} \in \mathbf{Dep}(Z)$ which in turn implies that Z and Y are tightly dependent.

Assumption 3 and Definition 2 sharpen Theorem 1 to what we define as a v-star. Denote the union of the minimal dependencies of Z by $\overline{\mathbf{Dep}}(Z) = \cup_{\mathcal{Y} \in \mathbf{Dep}(Z)} \mathcal{Y}$ and the union of the sets of the dependent members of the minimal dependencies by $\mathbf{Dep}^{\circ}(Z) = \cup_{\mathcal{Y} \in \mathbf{Dep}(Z)} \mathcal{Y}^{\circ}$.

Definition 3 (v-star) *A* v-star *over the node set $\mathcal{X}$ is the DAG where a single node $X \in \mathcal{X}$ is connected to every other node in $\mathcal{X}$ either directly (i.e., the two are adjacent) or indirectly by a collider whose effect node is also a node in $\mathcal{X}$. The DAG is also referred to as a* v-star centered at X.

Theorem 2 *Consider random variables $\mathcal{X}$ with joint probability distribution P, and let $Z \in \mathcal{X}$. If DAG $\mathcal{G}$ is an I-map for P and satisfies Assumption 3, then* (i) *Z and $\mathbf{Dep}^{\circ}(Z)$ form a v-star centered at Z; and* (ii) *for every $Y \in \mathbf{Dep}^{\circ}(Z)$ that is not adjacent to Z, a node in $\mathcal{Y} \setminus \{Y\}$ is a descendant of (or equals) the effect node of the collider connecting Y to Z, where $\mathcal{Y}$ is the minimal dependence of Z that includes Y.*

The structures in Examples 1 and 2 both appear to be v-stars. Theorem 2 does not reveal the structure of each individual minimal dependence but all of them as a whole: Dependent member Y of a minimal dependence $\mathcal{Y} \in \mathbf{Dep}(Z)$ may be connected to Z directly, indirectly by another node in $\mathcal{Y}$, or indirectly by another node in a different minimal dependence of Z, say $\mathcal{Y}'$. However, this requires the other node to belong to the minimal dependence of Y as otherwise, there is no link between them according to Assumption 3. So if none of the members of the minimal dependencies belong to any of the other minimal dependencies of Z, then Theorem 2 is further sharpened to each minimal dependence of Z forming a v-star. Another special case is for singleton minimal dependencies as illustrated by the following corollary.

Corollary 1 *Consider random variables $\mathcal{X}$ with joint probability distribution P, and let $Z \in \mathcal{X}$. If DAG $\mathcal{G}$ is an I-map for P and satisfies Assumption 3, then for every singleton strong minimal dependence $\{Y\} \in \mathbf{Dep}(Z)$ (that is, for every Y and Z where $Y \not\perp Z \mid \mathcal{U}$ for all $U \subseteq \mathcal{X} \setminus \{Y, Z\}$), it holds that Z is adjacent to Y.*

3.2 THE SYMMETRIC CASE

So far, Theorems 1 and 2 indicate the possibility of intertwined connections of the minimal dependencies in the true DAG. The connections can be partly separated upon the introduction of some symmetry to the minimal dependence.

Definition 4 (Symmetric minimal dependence) *Given variables $\mathcal{X}$ with joint probability distribution P, a* symmetric minimal dependence *is a set of random variables $\mathcal{Z} = \{Z_1, \ldots, Z_n\} \subseteq \mathcal{X}$, such that for every $i \in \{1, \ldots, n\}$, $\mathcal{Z}_{-i} \in \mathbf{Dep}(Z_i)$ and $\mathcal{Z}^{\circ}_{-i} = \mathcal{Z}_{-i}$.*

Namely, for each variable of a symmetric minimal dependence, the remainder of the set is its strong minimal dependence and all members of which are dependent members. The variable sets in both Examples 1 and 2 are symmetric minimal dependencies, but not those in Example 4. What structure does the true DAG of a symmetric minimal dependence take? It appears that in both examples, there is a node where all others are directing to. We show that this is generally true.

Definition 5 (Directed star) *A* directed star *over nodes $\mathcal{Z}$ consists of a single* center *node $Z \in \mathcal{X}$ that has an incoming edge from every other node $\mathcal{Z} \setminus \{Z\}$. The directed star is also referred to as a* star directed at *Z.*

Corollary 2 *Consider random variables $\mathcal{X}$ with joint probability distribution P. If DAG $\mathcal{G}$ is an I-map for P and satisfies Assumption 3, then the nodes of a symmetric minimal dependence admit a directed star in $\mathcal{G}$.*

Unlike the general results in Theorems 1 and 2 on the structure of the union of the minimal dependencies, here, the structure of a single minimal dependence is revealed. Namely, if a node X and its minimal dependence form a symmetric minimal dependence, their graphical representation includes a directed star, regardless of the rest of the nodes in $\overline{\mathbf{Dep}}(X)$. The center node, however, is undetermined. For example, for a symmetric minimal dependence of size three, the possible DAGs include all cases in Figure 2-a–c) as well as Figure 2-d–f) where an additional link is included. Thus, there can be several candidate DAGs for the true DAG, which are not distinguishable based on the distribution P. This motivates defining a class for such candidate DAGs as in the following section.

Remark 1 *A special case of a symmetric minimal dependence is a set of two variables X and Y, where $X \not\perp Y \mid \mathcal{U}$ for all $\mathcal{U} \in \mathcal{X} \setminus \{X, Y\}$. Then Corollary 2 implies that there is a link between X and Y in the true DAG. Therefore, as highlighted in Corollary 1, every node that forms a singleton minimal dependence of X is adjacent to X.*

3.3 MINIMAL DEPENDENCE EQUIVALENCE CLASS

The idea with PDAG class P-maps is to capture all DAGs that are an I-map for the distribution P, satisfy the faithfulness condition, and are indistinguishable with respect to the distribution P. Similarly, here, we can define a class of DAGs that all are an I-map for the distribution P and satisfy minimal dependence faithfulness, and that potentially any of them can be the true DAG.

So far, neither the definition of a minimal dependence nor the minimal dependence faithfulness assumption helps to determine the direction of the edges of the true DAG. This is because faithfulness does not hold and hence an independence between a triple of variables does not necessarily imply an immorality between them in the true DAG. More specifically, we build upon the following result for faithful distributions that helps in finding the orientation of undirected edges found in constraint-based algorithms [Koller and Friedman, 2009]. Suppose that in a P-map, nodes Y_1 and Y_2 are not adjacent but are both adjacent with X. Then

1. if Y_1, Y_2, X form the immorality $Y_1 \rightarrow X \leftarrow Y_2$ (that is Y_1 and Y_2 are not adjacent and both are linked to X), then $Y_1 \not\perp Y_2 \mid X, \mathcal{U}$ for all $\mathcal{U} \subseteq \mathcal{X} \setminus \{X, Y_1, Y_2\}$;
2. if Y_1, Y_2, X do not form an immorality, then $Y_1 \not\perp Y_2 \mid \mathcal{U}$ for all $\mathcal{U} \subseteq \mathcal{X} \setminus \{X, Y_1, Y_2\}$.

Now, in our case, instead of the single nodes Y_1 and Y_2, we have the minimal dependencies $\mathcal{Y}_1$ and $\mathcal{Y}_2$ that can be viewed as "super nodes" and that form directed stars centered at X. Accordingly, the above conditions are extended by replacing Y_i with $\mathcal{Y}_i$ and making necessary adjustments.

First, we extend the definition of immorality so that it also applies to node sets.

Definition 6 *Node X and node sets $\mathcal{Y}_1, \mathcal{Y}_2$ form an* immorality *if* (i) *there is no node in $\mathcal{Y}_1$ that is adjacent to a node in $\mathcal{Y}_2$ and* (ii) *$\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \{X\}$ forms a directed star centered at X.*

Assumption 4 (Minimal orientation) *Consider an SCM over variables $\mathcal{X}$, inducing DAG $\mathcal{G}$ and joint probability distribution P. For $X \in \mathcal{X}$, if $\mathcal{Y}_1 \cup \{X\}$ and $\mathcal{Y}_2 \cup \{X\}$ are symmetric minimal dependencies, where no node in $\mathcal{Y}_1$ is adjacent with a node in $\mathcal{Y}_2$, then*

1. *if $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \{X\}$ form an immorality in $\mathcal{G}$, then $\mathcal{Y}_1 \not\perp \mathcal{Y}_2 \mid X, \mathcal{U}$ for all $\mathcal{U} \subseteq \mathcal{X} \setminus (\{X\} \cup \mathcal{Y}_1 \cup \mathcal{Y}_2)$;*
2. *if $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \{X\}$ do not form an immorality in $\mathcal{G}$, then $\mathcal{Y}_1 \not\perp \mathcal{Y}_2 \mid \mathcal{U}$ for all $\mathcal{U} \subseteq \mathcal{X} \setminus (\{X\} \cup \mathcal{Y}_1 \cup \mathcal{Y}_2)$.*

The main use of the above assumption is the contra-position of the second case. For symmetric minimal dependencies $\mathcal{Y}_1 \cup \{X\}$ and $\mathcal{Y}_2 \cup \{X\}$ where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are not adjacent, if $\mathcal{Y}_1 \perp \mathcal{Y}_2 \mid \mathcal{U}$ for some $\mathcal{U} \subseteq \mathcal{X} \setminus (\{X\} \cup \mathcal{Y}_1 \cup \mathcal{Y}_2)$, then $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \{X\}$ form a directed start centered at X.

Definition 7 (MD-equivalence) *Given variables $\mathcal{X}$ with joint probability distribution P, the* MD-equivalence class *of P, denoted by $\mathcal{M}_{\mathrm{MD}}(P)$, is the set of every DAG $\mathcal{G}$ over $\mathcal{X}$ such that:*

1. *$\mathcal{G}$ is an I-map for P and satisfies Assumptions 3 and 4;*
2. *for every $Z \in \mathcal{X}$, the set $\mathbf{Dep}^{\mathrm{o}}(Z)$ admits a v-star centered at Z in $\mathcal{G}$.*

Any two DAGs in $\mathcal{M}_{\mathrm{MD}}(P)$ are said to be MD-equivalent with respect to P.

The six DAGs in Figure 2-a–f) are MD-equivalent. One could additionally require every symmetric minimal dependence to admit a directed star in $\mathcal{G}$ in order for it to belong to the MD-equivalence class; however, that would be an (indirect) implication of Condition ii) in Definition 7 as Corollary 2 is a result of Theorem 2. To simplify the representation of the MD-equivalence class, we define the following extended type of PDAGs. A *dashed PDAG* over vertices $\mathcal{X}$ is a tuple $(\mathcal{X}, \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3)$, where $\mathcal{E}_1 \in \mathcal{X} \times \mathcal{X}$ is the set of directed edges, and $\mathcal{E}_2, \mathcal{E}_3 \subseteq \{\{X_1, X_2\} \mid X_1, X_2 \in \mathcal{X}, X_1 \neq X_2\}$ are the set of undirected and dashed edges, respectively, and that no pair of nodes appears in more than one edge set (i.e., the directions in $\mathcal{E}_1$ ignored, the edge sets are mutually exclusive).

Definition 8 (Class dashed PDAG) *Let $\mathcal{M}_{\mathrm{MD}}(P)$ be the MD-equivalence class of a distribution P over variables $\mathcal{X}$. The* class dashed PDAG *of P is a dashed PDAG $\mathcal{H}$ over $\mathcal{X}$*

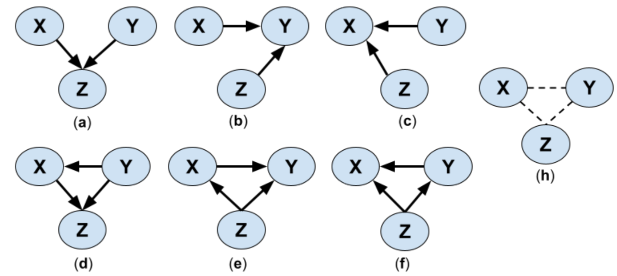

Figure 2: **a, b, c)** Three possible structures of an immorality over nodes X, Y, Z. **d, e, f)** Three possible structures of a triangle over nodes X, Y, Z. **h)** The graph representation for the existence of the immorality between nodes X, Y, Z.

satisfying the following conditions for every distinct pair $X, Y \in \mathcal{X}$:

1. *$\mathcal{H}$ contains no edge between X and Y if and only if X and Y are nonadjacent in every DAG in $\mathcal{M}_{\mathrm{MD}}(P)$;*
2. *$\mathcal{H}$ contains a directed edge $X \to Y$ if and only if $X \to Y$ occurs in every DAG in $\mathcal{M}_{\mathrm{MD}}(P)$;*
3. *$\mathcal{H}$ contains a solid undirected edge $X - Y$ if and only if X and Y are adjacent in every DAG in $\mathcal{M}_{\mathrm{MD}}(P)$, while both orientations occur among the DAGs in $\mathcal{M}_{\mathrm{MD}}(P)$;*
4. *$\mathcal{H}$ contains a dashed edge $X -- Y$ if and only if adjacency between X and Y is optional, namely, there exist $\mathcal{G}_1, \mathcal{G}_2 \in \mathcal{M}_{\mathrm{MD}}(P)$ such that X and Y are adjacent in $\mathcal{G}_1$ and nonadjacent in $\mathcal{G}_2$.*

A directed or solid undirected edge therefore represents an adjacency that is present in every DAG in the MD-equivalence class. In contrast, a dashed edge represents a possible adjacency that may be absent in some candidate DAGs.

The class dashed PDAG is interpreted together with the minimal dependences and symmetric minimal dependences detected by Algorithm 1. In a candidate DAG represented by the class dashed PDAG, every directed edge is retained with its displayed orientation, every solid undirected edge is assigned one of its two orientations, and every dashed edge may be deleted or retained with one of its two orientations. These choices must preserve acyclicity, must not introduce an immorality ruled out by the orientation stage, and must remain compatible with the v-star and directed-star structures implied by the detected minimal dependences. Thus, unlike a solid edge, a dashed edge does not assert adjacency in every candidate DAG.

3.4 RELATION TO OTHER TYPES OF FAITHFULNESS

The authors of [Ramsey et al., 2006] relaxed the faithfulness assumption to its implication in terms of the existence of a link. More specifically, they assume *adjacency faithfulness*, imposing that if nodes X and Y are adjacent in the true DAG $\mathcal{G}$, then they are dependent conditional on any subset of $\mathcal{X} \setminus \{X, Y\}$. Similarly, DAG $\mathcal{G}$ is said to be *minimally Markovian* if it is an I-map for P and two nodes are adjacent in $\mathcal{G}$ if and only if they are dependent conditioned on every subset of other variables [Sadeghi and Soo, 2022]. Both are the special case of Assumption 3 when all minimal dependencies are of size one. Moreover, the case where a symmetric minimal dependence is of size two matches the definition of triple unfaithfulness in [Marx et al., 2021] and hence is considered as a special case. In this regard, Assumptions 3 and 4 and Definition 1 and 2 can be considered as extensions of (2-)adjacency and (2-)orientation faithfulness and the notion of *k-association* in [Marx et al., 2021], respectively. Given all this, if a DAG is faithful, it is adjacency faithful. If a DAG is adjacency faithful, it is 2-adjacency faithful, and if it is 2-adjacency faithful it is minimal-dependence faithful. So minimal-dependence faithfulness can be considered as a generalization of the adjacency and 2-adjacency faithfulness, and in turn, faithfulness. This, however, comes with the inevitable cost that the set of final candidates for the true DAG increases. We highlight that the minimal dependence assumption imposed in this paper does not impose minimality on the number of edges and hence is different from existing notions such as *causal minimality* also known as *SGS minimality* stating that no proper sub-DAG of the true DAG $\mathcal{G}$ is an I-map for P [Spirtes et al., 2000, Zhang and Spirtes, 2008, Neapolitan, 2004]. It is also different from *P-minimality* [Zhang, 2013], stating that no DAG $\mathcal{G}'$ satisfying $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{G}')$ is an I-map for P. Another different notion is the *sparsest Markov representation (SMR)* which is an I-map $\mathcal{G}$, such that every other I-map that is not I-equivalent to $\mathcal{G}$ has more edges than $\mathcal{G}$ [Raskutti and Uhler, 2018a]. Similarly, an I-map $\mathcal{G}$ satisfying the *frugality* assumption, it is in the set of SMRs [Forster et al., 2018]. Overall, Assumption 3 does not imply any of the aforementioned minimality assumptions: the DAG in Figure 4-a) satisfies Assumption 3 but the sub-DAG in Figure 3-a) is the true DAG. Whether the aforementioned minimality assumptions imply Assumption 3 remains an open problem.

4 MD-PC ALGORITHM

How to perform causal discovery in the face of the unfaithfulness caused by a minimal dependence of size greater than one? A simple way is to iteratively find the $\mathbf{Dep}$ sets of all of the variables. However, the PC algorithm can be exploited to reduce the computations. To this end, we develop Algorithm 1, the *minimal dependence PC (MD-PC) algorithm*. Roughly speaking, the idea is to consider each minimal dependence as a super node and perform the PC algorithm as before. This determines the inter-structure of the minimal dependencies (super nodes). The intra-structure of the minimal dependencies are then partially determined by Theorem 2 and Corollary 2. This applies only to minimal dependencies of size at least two, resulting in an unfaithfulness, detected according to Proposition 1. The result is one class dashed PDAG representing a set of candidate DAGs, and this represented class contains the true DAG under Assumptions Assumption 3 and Assumption 4. Similar to PC, the algorithm goes through every pair of nodes X and Y. If the two are dependent, then Y does not belong to a minimal dependence of X of size greater than one, and hence, normal PC is followed. Namely, if later, the two become conditionally independent, the connecting link is deleted; otherwise, Y is adjacent with X (equivalently, $\{Y\} \in \mathbf{Dep}(X)$). However, if X and Y are independent, we investigate the possibility of Y forming a strong minimal dependence of size at least two with some other nodes. If yes, the minimal dependence is stored and all links between X and the minimal dependence are removed. Otherwise, the link between X and Y is removed. Next, we proceed to mark the minimal dependencies using the minimal orientation rule.

Theorem 3 *Consider an SCM over variables $\mathcal{X}$, inducing DAG $\mathcal{G}$ and joint probability distribution P. If $\mathcal{G}$ satisfies Assumptions 3 and 4, then Algorithm 1 outputs the MD-equivalent class dashed PDAG of P.*

Remark 2. Compared to PC, the MD-PC algorithm also computes $\mathbf{Dep}(X)$ for all X, which can add computational burden despite leveraging PC's results. Both algorithms have a complexity of $\mathcal{O}(2^N)$, but for sparse DAGs, PC often requires fewer operations. The MD-PC algorithm may not benefit similarly, as its complexity depends on variable dependencies. If p is the maximum number of parents in the true DAG, the number of CI tests remains bounded by $\mathcal{O}(N^p)$, similar to PC. Thus, knowing p can help mitigate MD-PC's complexity in sparse high-dimensional networks.

4.1 ORACLE AND FINITE-SAMPLE EVALUATION

Oracle evaluation. We evaluated MD-PC using exact conditional-independence information. On the generalized XOR model, Example 2, and Appendix Example 5, oracle MD-PC recovered the represented class implied by the corresponding exact strong minimal dependences. In each case, the represented class contains the generating DAG. In contrast, PC removes the true MD adjacencies between X and the variables Y_i because $X \perp Y_i$.

Finite-sample setting. We next considered deterministic and noisy XOR models, $X = Y_1 \oplus \cdots \oplus Y_q,\ X = Y_1 \oplus$

Table 1: Finite-sample results for the common deterministic-parity setting. MD-edge recall counts both solid and dashed possible edges as recovered.

Method	MD-edge recall	MD-set recall	CI tests/run
PC-stable	0.011	0.000	6.5
CPC	0.011	0.000	6.5
GS Markov blanket	0.017	0.000	6.8
HillClimb/BDeu	0.006	0.000	N/A
MD-PC, bounded sample	0.728	0.635	24.3

$\cdots \oplus Y_q \oplus E$, where $Y_1, \ldots, Y_q$ are independent Bernoulli variables and $E \sim \mathrm{Bernoulli}(\epsilon)$ is independent output noise. We tested the conditional-independence-based methods, PC-stable Colombo and Maathuis [2014], CPC [Ramsey et al., 2006], Grow-Shrink Margaritis and Thrun [1999b] Markov blanket, and bounded sample MD-PC. We used the same discrete conditional-independence test with nominal level $\alpha = 0.01$.

For node X, the ground-truth MD set is $\mathcal{Y}_X^\star = Y_1, \ldots, Y_q$, for $q \in \{2, 3, 4\}$. A true MD adjacency between X and Y_i is counted as recovered when the output contains either a solid edge (directed or undirected) or a dashed possible edge between X and Y_i. Exact MD-set recall counts a run as successful only when the complete ground-truth set $\mathcal{Y}_X^\star$ is returned as an estimated MD set for X; a proper subset is not counted as a recovery. All reported values are averaged over $R = 5$ independent repetitions.

Table 1 reports results for the common deterministic-parity setting. PC-stable, CPC, GS Markov blanket, and HillClimb/BDeu recover almost none of the true MD adjacencies. In contrast, bounded sample MD-PC obtains substantially higher MD-edge and exact MD-set recall, at the cost of additional conditional-independence tests. In the noisy-parity experiments, bounded sample MD-PC attained at least 80% exact MD-set recall in the tested sample-size range.

5 CONCLUSION

We studied causal structure learning when a variable can be independent of each individual member of a set while depending on that set jointly. We formalized this pattern through minimal dependence and introduced minimal-dependence faithfulness and minimal orientation as relaxations of standard faithfulness. Under the I-map condition, Assumptions 3–4, and oracle conditional-independence information, Theorem 3 shows that MD-PC returns the class dashed PDAG representing the corresponding MD-equivalence class. The finite-sample experiments demonstrate that bounded sample MD-PC can detect parity-type minimal dependences that are largely missed by the PC-style and score-based baselines considered here. These experiments are empirical evidence for detectability in this model family; they do not establish a general finite-sample recovery guarantee, a general scalability guarantee, or superiority on arbitrary unfaithful distributions. Extending the analysis to broader forms of unfaithfulness and obtaining finite-sample guarantees are important directions for future work.

References

H. Andersen. When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, 80(5): 672–683, 2013.

Diego Colombo and Marloes H. Maathuis. Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, 15(116):3921–3962, 2014.

Robert Culverhouse, Brian K. Suarez, Jennifer Lin, and Theodore Reich. A perspective on epistasis: Limits of models displaying no main effect. *The American Journal of Human Genetics*, 70(2):461–471, 2002.

Ron Eccles. Understanding the symptoms of the common cold and influenza. *The Lancet infectious diseases*, 5(11): 718–725, 2005.

M. Forster, G. Raskutti, R. Stern, and N. Weinberger. The frugal inference of causal relations. *The British Journal for the Philosophy of Science*, 2018.

Scott B Halstead. Dengue. *The lancet*, 370(9599):1644–1652, 2007.

S. Horii. Bayesian model averaging for causality estimation and its approximation based on gaussian scale mixture distributions. In *International Conference on Artificial Intelligence and Statistics*, pages 955–963, 2021.

Hamid Kalantari, Russell Greiner, and Pouria Ramazi. Extendable and iterative structure learning strategy for bayesian networks. In *International Conference on Learning Representations*, 2025.

N.R. Ke, O. Bilaniuk, A. Goyal, S. Bauer, H. Larochelle, B. Schölkopf, M.C. Mozer, C. Pal, and Y. Bengio. Learning neural causal models from unknown interventions. 2020.

N.K. Kitson, A.C. Constantinou, Z. Guo, Y. Liu, and K. Chobtham. A survey of bayesian network structure learning. 2021.

Murat Kocaoglu. Characterization and learning of causal graphs with small conditioning sets. In *Advances in Neural Information Processing Systems*, volume 36, pages 74140–74179, 2023.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

J. Lemeire, S. Meganck, F. Cartella, and T. Liu. Conservative independence-based causal structure learning in absence of adjacency faithfulness. *International Journal of Approximate Reasoning*, 53(9):1305–1325, 2012.

D. Margaritis and S. Thrun. Bayesian network induction via local neighborhoods. In *Advances in Neural Information Processing Systems*, volume 12, 1999a.

Dimitris Margaritis and Sebastian Thrun. Bayesian network induction via local neighborhoods. In Sara A. Solla, Todd K. Leen, and Klaus-Robert Müller, editors, *Advances in Neural Information Processing Systems 12*, pages 505–511. MIT Press, 1999b.

A. Marx, A. Gretton, and J.M. Mooij. A weaker faithfulness assumption based on triple interactions. In *Uncertainty in Artificial Intelligence*, pages 451–460, 2021.

Alison A. Motsinger and Marylyn D. Ritchie. Multifactor dimensionality reduction: An analysis strategy for modelling and detecting gene–gene interactions in human genetics and pharmacogenomics studies. *Human Genomics*, 2(5):318–328, 2006.

R.E. Neapolitan. *Learning Bayesian Networks*, volume 38. Pearson Prentice Hall, 2004.

I. Ng, Y. Zheng, J. Zhang, and K. Zhang. Reliable causal discovery with improved exact search and weaker assumptions. In *Advances in Neural Information Processing Systems*, volume 34, pages 20308–20320, 2021.

J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017.

J. Ramsey, J. Zhang, and P.L. Spirtes. Adjacency-faithfulness and conservative causal inference. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 401–408, 2006.

J. Ramsey, J. Zhang, and P.L. Spirtes. Adjacency-faithfulness and conservative causal inference. *arXiv preprint arXiv:1206.6843*, 2012.

G. Raskutti and C. Uhler. Learning directed acyclic graph models based on sparsest permutations. *Stat*, 7(1):e183, 2018a.

Garvesh Raskutti and Caroline Uhler. Learning directed acyclic graph models based on sparsest permutations. *Stat*, 7(1):e183, 2018b.

Marylyn D. Ritchie, Lance W. Hahn, Nady Roodi, L. Renee Bailey, William D. Dupont, Fritz F. Parl, and Jason H. Moore. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *The American Journal of Human Genetics*, 69(1):138–147, 2001.

Mohammad Hossein Roohi, Pouria Ramazi, and Tongwen Chen. Towards accurate root-alarm identification: The causal bayesian network approach. In *2021 5th International Conference on Control and Fault-Tolerant Systems (SysTol)*, pages 169–174. IEEE, 2021.

K. Sadeghi and T. Soo. Conditions and assumptions for constraint-based causal structure learning. *Journal of Machine Learning Research*, 23(109):1–34, 2022.

Kayvan Sadeghi. Faithfulness of probability distributions and graphs. *Journal of Machine Learning Research*, 18 (148):1–29, 2017.

Vahideh Sanjaroonpouri and Pouria Ramazi. Linear scm identification in the presence of confounders and gaussian noise. In *International Conference on Learning Representations*, 2025.

Jinxing Shang, Jinxing Liu, and Boxin Guan. EpiReSIM: A resampling method of epistatic model without marginal effects using under-determined system of equations. *Genes*, 13(12):2286, 2022.

P. Spirtes and J. Zhang. A uniformly consistent estimator of causal effects under the k-triangle-faithfulness assumption. *Statistical Science*, pages 662–678, 2014.

P. Spirtes, C.N. Glymour, R. Scheines, and D. Heckerman. *Causation, Prediction, and Search*. MIT Press, 2000.

Marcel Wienöbst and Maciej Liskiewicz. Recovering causal structures from low-order conditional independencies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9690–9697, 2020.

J. Zhang. A comparison of three occam's razors for markovian causal models. *The British Journal for the Philosophy of Science*, 2013.

J. Zhang and P. Spirtes. Detection of unfaithfulness and robust causal inference. *Minds and Machines*, 18(2): 239–271, 2008.

J. Zhang and P. Spirtes. The three faces of faithfulness. *Synthese*, 193:1011–1027, 2016.

Proofs, Examples and Algorithm (Supplementary Material)

Pouria Ramazi[1] **Hamid Kalantari**[2]

[1]Departments of Mathematics and Statistics and Biological Sciences, University of Calgary, Calgary, Alberta, Canada
[2]Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

A APPENDIX

Definition 9 (d-separation) *[Ramsey et al., 2012] Consider DAG $\mathcal{G}$ with node set $\mathcal{X}$. A trail $\mathcal{T}$ between two nodes $X, Y \in \mathcal{X}$ is* active *relative to (or given) a set of nodes $\mathcal{Z} \subseteq \mathcal{X}$ if* (i) *for each collider on $\mathcal{T}$, at least one of the descendants of the collider node is in $\mathcal{Z}$, and* (ii) *no other node on $\mathcal{T}$ is in $\mathcal{Z}$. The node subsets $\mathcal{X}_1, \mathcal{X}_2 \subseteq \mathcal{X}$ are* d-separated *given $\mathcal{Z}$, denoted $\mathrm{d} - sep_{\mathcal{G}}(\mathcal{X}_1, \mathcal{X}_2 \mid \mathcal{Z})$, if there is no active trail between any node $X_1 \in \mathcal{X}_1$ and any node $X_2 \in \mathcal{X}_2$ given $\mathcal{Z}$. The set of all d-separations in $\mathcal{G}$ is denoted by $\mathcal{I}(\mathcal{G})$.*

Example 4 (Non-Bernoulli distribution; non-deterministic) *The following is a V-structure SCM where the collider node X does not have a deterministic relationship with Y_1 and Y_2 and does not take a Bernoulli distribution. The SCM consists of the variables X, Y_1, Y_2 and is defined by the equations*

$$\begin{cases} X := f(Y_1, Y_2, U) \\ Y_1 := \mathrm{Bernoulli}(1/2) \\ Y_2 := \mathrm{Bernoulli}(1/4) \end{cases}$$

where $U \in \{1, 2, 3, 4, 5\}$ is the noise variable with the following distribution

$$P(U) = \begin{cases} .2 & U = 1 \\ .1 & U = 2 \\ .1 & U = 3 \\ .1 & U = 4 \\ .5 & U = 5 \end{cases}$$

and the function f is defined as follows:

$$f(0,1,U) = \begin{cases} 0 & U = 1 \\ 1 & U > 1 \end{cases}, f(1,0,U) = \begin{cases} 0 & U \leq 2 \\ 1 & U > 2 \end{cases},$$

$$f(1,1,U) = \begin{cases} 0 & U \leq 3 \\ 1 & U > 3 \end{cases}, f(0,0,U) = \begin{cases} 0 & U = 4 \\ 1 & U > 4 \end{cases}$$

It can be verified that $\mathcal{Y} = \{Y_1, Y_2\}$ is the minimal dependence of X.

Example 5 (Non-Bernoulli distribution; non-deterministic) *The following is another V-structure SCM with binary nodes $X, Y_1 \in \{0, 1\}$ and non-binary node $Y_2 \in \{0, 1, 2\}$:*

$$\begin{cases} X := f(Y_1, Y_2, U) \\ Y_1 := \mathrm{Bernoulli}(0.4) \\ Y_2 := \mathrm{Multinomial}(0.6, 0.2, 0.2) \end{cases}$$

where for brevity we only provide the conditional probability distribution of X:

$$P(X = 0 \mid Y_1, Y_2) = \begin{cases} 0.65 & Y_1 = 0, Y_2 = 0 \\ 0.25 & Y_1 = 0, Y_2 = 1 \\ 0.1 & Y_1 = 0, Y_2 = 2 \\ 0.33 & Y_1 = 1, Y_2 = 0 \\ 0.6 & Y_1 = 1, Y_2 = 1 \\ 0.7 & Y_1 = 1, Y_2 = 2 \end{cases}$$

Again, it can be verified that $\mathcal{Y} = \{Y_1, Y_2\}$ is the minimal dependence of X.

Example 6 (Water tank) *Consider a water tank that has an input flow and an output flow (equal flow rate) controlled by two valves. Let $X \in \{0, 1\}$ denote the water level of the tank, which is 0 if the water level equals some desired level, e.g., 1 meter above the base, and is 1 otherwise. Also, denote by $Y_1 \in \{0, 1\}$ and $Y_2 \in \{0, 1\}$ the valve statuses of the input and output flows, where a value of 0 means that the valve is open and 1 means that it is closed. Clearly, whether the water level changes from the desired level (X) is independent of just Y_1 (the inflow tap being open or closed) or just Y_2 (the output flow valve being open or closed), but is not independent of any of them once the other one is known. For example, given that the inflow valve is closed, then the output flow valve being open implies that the water level decreases, and being closed implies that the water level remains unchanged. It follows that $\mathcal{Y} = \{Y_1, Y_2\}$ is a minimal dependence for X.*

Proof of Proposition 1. We prove by contradiction. Assume that for some $X \in \mathcal{X}$, there exists a minimal dependence $\mathcal{Y}$ of X where $|\mathcal{Y}| \geq 2$, and assume on the contrary that there exists some DAG $\mathcal{G}$ to which P is faithful. In view of faithfulness, Condition (2) yields the absence of any active trail between X and each node in $\mathcal{Y}$. We show that X is d-separated from every node $Y_i \in \mathcal{Y}$ given $\mathcal{Y}_{-i}$. Otherwise, for some $i \in \{1, \dots, n\}$, there must be an inactive trail between X and Y_i that becomes active upon the observation of $\mathcal{Y}_{-i}$. The only way this is possible is to have on the inactive trail including one or more colliders where each has a node in $\mathcal{Y}_{-i}$ as its descendant (or effect node). Denote by V_1 the closest node to X on this trail that together with two other nodes V_2 and V_3 on this trail form the collider $V_1 \to V_2 \leftarrow V_3$ where some node $Y_k \in \mathcal{Y}_{-i}$ is a descendant of (or equals) V_2. Let $\mathcal{T}_1$ be the directed path from X to V_2 on this trail and $\mathcal{T}_2$ be a directed path from V_2 to Y_k Then the concatenated path $\mathcal{T}_1\mathcal{T}_2$ is a collider-free and hence active trail from X to a node in $\mathcal{Y}_{-i}$ and in turn $\mathcal{Y}$, a contradiction. Hence, for all $i \in \{1, \dots, n\}$, $X \perp Y_i \mid \mathcal{Y}_{-i}$ belongs to $\mathcal{I}(\mathcal{G})$ but not to $\mathcal{I}(P)$ according to (1). This violates the I-map assumption, a contradiction. □

Proof of Proposition 2. Eq. 6 follows from the third condition in Definition 1. To prove (5), let $\mathcal{Y}' = \{Y_1', \dots, Y_k'\}, k < n$. Then in view of (6), $P(X \mid \mathcal{Y}') = P(X \mid \mathcal{Y}' \setminus \{Y_1'\})$. Next, by letting $\mathcal{Y}' \setminus \{Y_1'\}$ to be $\mathcal{Y}'$ in (6), and Y to be Y_2', we obtain $P(X \mid \mathcal{Y}' \setminus \{Y_1'\}) = P(X \mid \mathcal{Y}' \setminus \{Y_1', Y_2'\})$. Hence, by induction, it can be shown that $P(X \mid \mathcal{Y}') = P(X)$. To prove (3),let $\mathcal{Y} = \{Y_1, \dots, Y_n\}$ and assume that (1) holds for $Y = Y_1$. We prove by contradiction. Assume on the contrary that (1) does not hold for some $Y \in \mathcal{Y}$, say Y_2, i.e., $X \perp Y_2 \mid \mathcal{Y} \setminus \{Y_2\}$. Then $P(X \mid \mathcal{Y}) = P(X \mid \mathcal{Y} \setminus \{Y_2\})$. Now, by letting $\mathcal{Y}' = \mathcal{Y} \setminus \{Y_2\}$, (5) results in $P(X \mid \mathcal{Y} \setminus \{Y_2\}) = P(X\})$. Thus, $P(X \mid \mathcal{Y}) = P(X)$. On the other hand, again from (5) , $P(X \mid \mathcal{Y} \setminus \{Y_1\}) = P(X)$. Hence, $P(X \mid \mathcal{Y}) = P(X \mid \mathcal{Y} \setminus \{Y_1\})$, yielding $X \perp Y_1 \mid \mathcal{Y} \setminus \{Y_1\}$, a contradiction, completing the proof. Finally, (4) is also proven by contradiction as if on the contrary $X \perp \mathcal{Y}$, then $P(X \mid \mathcal{Y}) = P(X)$ which similar to the proof for (5) results in a contradiction. □

Proof of Theorem 1. Consider an arbitrary $Y_i \in \mathcal{Y}$. Should Y_i be adjacent to Z the result is trivial, so consider otherwise. In view of Condition (1), $Z \not\perp Y_i \mid \mathcal{Y}_{-i}$. So $\mathcal{Y}_{-i}$ observed, $\mathcal{G}$ being an I-map implies an active trail, say $\mathcal{T}$, between Z and Y_i. Thus, every collider on trail $\mathcal{T}$ becomes active upon observing $\mathcal{Y}_{-i}$, implying that the effect node or its descendant is in $\mathcal{Y}_{-i}$, and hence, $\mathcal{Y}$. □

Proof of Theorem 2. Consider an arbitrary dependent member $Y \in \mathbf{Dep}^{\circ}(Z)$. Let $\mathcal{Y} \in \mathbf{Dep}(Z)$ be a minimal dependence containing Y. If Y is adjacent to Z, the result is immediate. Suppose instead that Y is not adjacent to Z. Take $\mathcal{U} = \mathcal{X} \setminus (Z \cup \mathcal{Y})$. Since Y is a dependent member of $\mathcal{Y}$, Definition 2 gives $Z \not\perp_P Y \mid (\mathcal{Y} \setminus Y) \cup \mathcal{U} = \mathcal{X} \setminus Z, Y$. Because $\mathcal{G}$ is an I-map for P, there is an active trail between Z and Y given $\mathcal{X} \setminus Z, Y$. Every internal node of this trail is conditioned on. Hence, no internal node can be a non-collider, since such a node would block the trail. A trail with more than one internal node cannot have every internal node as a collider, because two consecutive internal nodes cannot both receive arrowheads from their common edge. Therefore, the active trail must have the form $Z \to W \leftarrow Y$. Since Z and W are adjacent, Assumption 3 implies that $W \in \mathbf{Dep}^{\circ}(Z)$. Thus, every dependent member of Z is either adjacent to Z or connected to Z through a collider whose effect node is in $\mathbf{Dep}^{\circ}(Z)$. This proves part (i). For part (ii), apply Definition 2 with $\mathcal{U} = \varnothing$. The dependence between

Z and Y then holds given $\mathcal{Y} \setminus Y$. Since $Z \perp_P Y$ by condition (8), this dependence must be activated by conditioning on $\mathcal{Y} \setminus Y$. Hence, by the active-trail argument used in the proof of Theorem 1, an activating collider can be chosen whose effect node is either a member of $\mathcal{Y} \setminus Y$ or has a descendant in $\mathcal{Y} \setminus Y$. This proves part (ii). □

Proof of Corollary 1. The proof follows Part *(ii)* in Theorem 2 and the fact that if a minimal dependence $\mathcal{Y} \in \mathbf{Dep}(Z)$ is a singleton, say $\{Y\}$, then $\mathcal{Y} \setminus \{Y\}$ becomes empty, implying that a collider cannot connect Y and Z. So the two are adjacent. □

Proof of Corollary 2. Let $\mathcal{Z} = Z_1, \ldots, Z_n$ be a symmetric minimal dependence. Define a directed reachability relation on $\mathcal{Z}$ by $Z_i \preceq Z_j$ if and only if $Z_i = Z_j$ or there is a directed path from Z_i to Z_j in $\mathcal{G}$. Since $\mathcal{G}$ is acyclic, this relation has a maximal element. Denote one such element by $Z^\star$. Thus, there is no directed path from $Z^\star$ to any other node in $\mathcal{Z}$. We show that every $Z_i \in \mathcal{Z} \setminus Z^\star$ is adjacent to $Z^\star$. Suppose otherwise. Applying Theorem 2 to the pair $Z^\star$ and Z_i within the symmetric minimal dependence gives a collider $Z^\star \to W \leftarrow Z_i$, where either $W \in \mathcal{Z} \setminus Z^\star, Z_i$, or W has a descendant $Z_j \in \mathcal{Z} \setminus Z^\star, Z_i$. In the first case, $Z^\star \to W$ is a directed path from $Z^\star$ to another node in $\mathcal{Z}$. In the second case, $Z^\star \to W \to \cdots \to Z_j$ is a directed path from $Z^\star$ to another node in $\mathcal{Z}$. Both cases contradict the maximality of $Z^\star$. Hence, every node in $\mathcal{Z} \setminus Z^\star$ is adjacent to $Z^\star$. Finally, none of these adjacencies can be oriented $Z^\star \to Z_i$, because that would create a directed path from $Z^\star$ to another member of $\mathcal{Z}$. Therefore, every such edge is oriented $Z_i \to Z^\star$ for each $Z_i \in \mathcal{Z} \setminus Z^\star$. Thus, the nodes of $\mathcal{Z}$ admit a directed star centered at $Z^\star$. □

Proof of Theorem 3. We prove the oracle version of Algorithm 1, namely that every conditional-independence query is answered correctly. Let $G^\star$ denote the true DAG. Assume that $G^\star$ is an I-map for P and satisfies Assumptions 3 and 4.

$$\mathbf{Dep}^{\circ}(\mathcal{Y}, X).$$

The hatted versions of these objects denote the quantities computed by Algorithm 1.

Step 1: Detection of non-singleton MD sets. Fix $X \in \mathcal{X}$. Every member of a non-singleton minimal dependence $\mathcal{Y} \in \mathcal{D}_{\geq 2}(X)$ belongs to $\mathrm{Perp}(X)$ because Definition 2 gives

$$X \perp_P Y, \qquad Y \in \mathcal{Y}.$$

Therefore, every true non-singleton MD set is considered during the MD-detection stage.

Suppose that $\mathcal{Y} \in \mathcal{D}_{\geq 2}(X)$. Candidates are examined in increasing cardinality, and a candidate is skipped only if it strictly contains an already detected MD set. Thus, no proper subset of $\mathcal{Y}$ can prevent $\mathcal{Y}$ from being tested. For every dependent member $Y \in \mathcal{D}^{\circ}(\mathcal{Y}, X)$, Definition 2 gives

$$\mathsf{SD}(X, Y; \mathcal{Y} \setminus Y).$$

Hence, $\mathcal{Y}$ is accepted.

$$Y \in \mathcal{C} : \mathsf{SD}(X, Y; \mathcal{C} \setminus Y),$$

which is exactly $\mathcal{D}^{\circ}(\mathcal{C}, X)$.

Step 2: Temporary removal of non-singleton MD pairs. The removal of a solid edge during the MD-detection stage is not a final non-adjacency decision. It only removes that pair from the temporary ordinary skeleton because the relation has already been stored as a non-singleton MD relation.

If

$$Y \notin \mathbf{Dep}^{\circ}(X),$$

then Assumption 3 excludes an adjacency between X and Y in any admissible DAG. If instead $Y \in \mathbf{Dep}^{\circ}(X)$, the relation is retained in the stored MD information and is later represented by a compelled directed edge or by a dashed edge. Thus, the temporary removal does not discard a possible MD adjacency.

Step 3: Recovery of the residual singleton skeleton. Consider a true singleton-dependence adjacency $X - Y$. Then $Y \in \mathbf{Dep}^{\circ}(X)$, and Definition 2 gives

$$X \not\perp_P Y \mid \mathcal{U}$$

for every

$$\mathcal{U} \subseteq \mathcal{X} \setminus X, Y.$$

Hence, no oracle conditional-independence query in the PC-style skeleton stage can remove the solid edge $X - Y$.

Conversely, let X and Y be nonadjacent in $G^\star$ and suppose that their relation has not already been stored as a non-singleton MD relation. By the standard local-separation property of DAGs, there is a d-separating set $\mathcal{U}$ contained in the true adjacency set of one endpoint. Every true neighbor of X is either a singleton-dependence neighbor, in which case it remains a solid neighbor, or a dependent member of a non-singleton MD set, in which case it belongs to $\widehat{\mathbf{Dep}^{\mathrm{o}}} \geq 2(X)$. Therefore,

$$\mathcal{U} \subseteq \operatorname{Adj} \mathrm{s}(\mathcal{H}, X) \cup \widehat{\mathbf{Dep}^{\mathrm{o}}}_{\geq 2}(X)$$

at the relevant stage of the search. Since $G^\star$ is an I-map for P, this d-separation implies

$$X \perp_P Y \mid \mathcal{U}.$$

The PC-style stage eventually tests $\mathcal{U}$ and removes the solid edge. Therefore, the remaining solid skeleton is exactly the residual singleton-dependence skeleton.

Step 4: Singleton and symmetric MD sets. Every remaining solid neighbor Y of X after the skeleton stage corresponds to a singleton minimal dependence Y. Thus, the singleton-MD stage adds exactly the singleton minimal dependences and their dependent members.

After the non-singleton and singleton stages, all minimal dependences and dependent members have been identified. The symmetric-MD stage then applies Definition 4 directly. Hence, Sym contains exactly the symmetric minimal dependences.

Step 5: Orientation of compelled MD immoralities. Consider two distinct minimal dependences $\mathcal{C}_1, \mathcal{C}_2 \in \mathbf{Dep}(X)$ such that $\mathcal{C}_1 \cup X$ and $\mathcal{C}_2 \cup X$ are symmetric and no member of $\mathcal{C}_1$ is adjacent to a member of $\mathcal{C}_2$.

The orientation stage directs the corresponding edges toward X only when there exists a common set

$$\mathcal{U} \subseteq \mathcal{X} \setminus (X \cup \mathcal{C}_1 \cup \mathcal{C}_2)$$

such that

$$\mathcal{C}_1 \perp_P \mathcal{C}_2 \mid \mathcal{U}.$$

If the true structure were not an immorality centered at X, Assumption 4(ii) would imply dependence for every such $\mathcal{U}$, contradicting this criterion. Hence, every orientation made at this stage is sound.

Conversely, for a compelled MD immorality centered at X, Assumption 4(i) implies that a separating set detected for the two MD sets cannot contain X. Thus, the common separator satisfies the orientation criterion and the corresponding edges are directed toward X. The orientation stage is therefore complete for compelled MD immoralities.

Step 6: Dashed relations and closure of the solid subgraph. For every non-singleton MD set $\mathcal{C} \in \mathbf{Dep}_{\geq 2}(X)$ and every dependent member

$$Y \in \mathbf{Dep}^{\mathrm{o}}(\mathcal{C}, X),$$

the algorithm adds a dashed edge between X and Y unless that relation has already been compelled to be directed. This is sound because pairs outside $\mathbf{Dep}^{\mathrm{o}}(X)$ cannot be adjacent to X by Assumption 3. It is complete because every dependent member of every non-singleton MD set was identified in Step 1.

Finally, the closure stage is applied only to the solid subgraph. Dashed edges are not treated as ordinary adjacencies and do not participate in acyclicity or no-new-immorality propagation. Every orientation made in this stage is therefore sound because the opposite orientation would create a directed cycle or an unshielded collider ruled out by the stored separating information.

Combining Steps 1–6, the output has the directed, solid undirected, absent, and dashed edge semantics of Definition 8. Therefore, Algorithm 1 outputs the class dashed PDAG representing the MD-equivalence class of P, and this class contains the true DAG $G^\star$. $\square$

Algorithm: Notations and Definitions. Let $\mathsf{SD}(X, Y; \mathcal{V})$ denote the strong-dependence condition

$$X \not\perp_P Y \mid \mathcal{V} \cup \mathcal{U} \quad \text{for all} \quad \mathcal{U} \subseteq \mathcal{X} \setminus (\{X, Y\} \cup \mathcal{V}).$$

Also, let $\mathsf{Csep}_X(\mathcal{C}_1, \mathcal{C}_2)$ mean that there exists

$$\mathcal{U} \subseteq \mathcal{X} \setminus (\{X\} \cup \mathcal{C}_1 \cup \mathcal{C}_2)$$

such that $\mathcal{C}_1 \perp_P \mathcal{C}_2 \mid \mathcal{U}$.

In Algorithm 1, an eligible triple $(X, \mathcal{C}_1, \mathcal{C}_2)$ satisfies

$$\mathcal{C}_1 \cup \{X\}, \mathcal{C}_2 \cup \{X\} \in \mathrm{Sym},$$

and no node in $\mathcal{C}_1$ is adjacent to a node in $\mathcal{C}_2$.

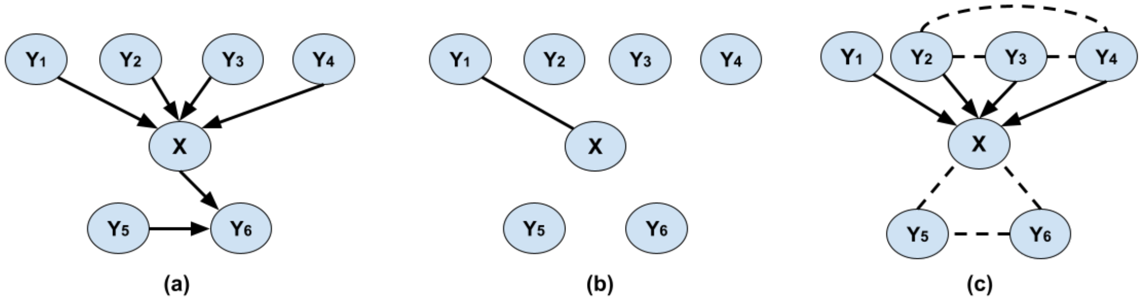

Figure 3: **a)** The true DAG. **b)** Output of the PC algorithm. **c)** The dashed PDAG representing the MD-equivalent class from the MD-PC algorithm.

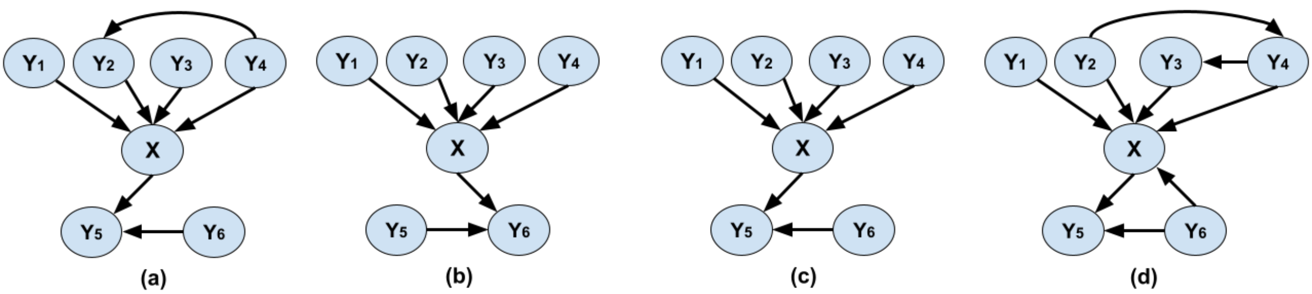

Figure 4: Examples of the MD-equivalent class.

Algorithm 1: Oracle MD-PC Algorithm

Input: $\mathcal{X}$ and oracle CI information for P
Output: $(\mathcal{H}, \widehat{\mathrm{Dep}}, \widehat{\mathbf{Dep}^{\circ}}, \mathrm{Sym})$
Initialize $\mathcal{H}$ as the complete solid undirected graph; $\mathrm{Sym} \leftarrow \varnothing$
foreach $X \in \mathcal{X}$ **do**
  $\mathrm{Perp}(X) \leftarrow \{Y \in \mathcal{X} \setminus \{X\} : X \perp_P Y\}$; $\widehat{\mathrm{Dep}}_{\geq 2}(X) \leftarrow \varnothing$
foreach $X \in \mathcal{X}$ **do**
  foreach $\mathcal{C} \subseteq \mathrm{Perp}(X)$, $|\mathcal{C}| \geq 2$, *in increasing* $|\mathcal{C}|$ **do**
    if $\nexists \mathcal{D} \in \widehat{\mathrm{Dep}}_{\geq 2}(X)$ *with* $\mathcal{D} \subsetneq \mathcal{C}$ **then**
      $\mathcal{C}^{\circ} \leftarrow \{Y \in \mathcal{C} : \mathsf{SD}(X, Y; \mathcal{C} \setminus \{Y\})\}$
      if $\mathcal{C}^{\circ} \neq \varnothing$ **then**
        $\widehat{\mathrm{Dep}}_{\geq 2}(X) \leftarrow \widehat{\mathrm{Dep}}_{\geq 2}(X) \cup \{\mathcal{C}\}$; $\widehat{\mathbf{Dep}^{\circ}}(\mathcal{C}, X) \leftarrow \mathcal{C}^{\circ}$
  $\widehat{\mathbf{Dep}^{\circ}}_{\geq 2}(X) \leftarrow \bigcup_{\mathcal{C} \in \widehat{\mathrm{Dep}}_{\geq 2}(X)} \widehat{\mathbf{Dep}^{\circ}}(\mathcal{C}, X)$
  Remove every solid edge $X - Y$ with $Y \in \mathcal{C}$ and $\mathcal{C} \in \widehat{\mathrm{Dep}}_{\geq 2}(X)$
$\mathbf{Sepset}(X, Y) \leftarrow \varnothing$ for all $X, Y \in \mathcal{X}$; $m \leftarrow 0$
while $\max_X |\mathrm{Adj}_{\mathrm{s}}(\mathcal{H}, X) \cup \widehat{\mathbf{Dep}^{\circ}}_{\geq 2}(X)| > m$ **do**
  foreach *solid edge* $X - Y$ *in* $\mathcal{H}$ **do**
    foreach $\mathcal{U} \subseteq [\mathrm{Adj}_{\mathrm{s}}(\mathcal{H}, X) \cup \widehat{\mathbf{Dep}^{\circ}}_{\geq 2}(X)] \setminus \{Y\}$ *with* $|\mathcal{U}| = m$ **do**
      if $X \perp_P Y \mid \mathcal{U}$ **then**
        Remove $X - Y$; $\mathbf{Sepset}(X, Y) \leftarrow \mathbf{Sepset}(Y, X) \leftarrow \mathcal{U}$; **break**
  $m \leftarrow m + 1$
foreach $X \in \mathcal{X}$ **do**
  $\widehat{\mathrm{Dep}}(X) \leftarrow \widehat{\mathrm{Dep}}_{\geq 2}(X) \cup \{\{Y\} : Y \in \mathrm{Adj}_{\mathrm{s}}(\mathcal{H}, X)\}$
  $\widehat{\mathbf{Dep}^{\circ}}(\{Y\}, X) \leftarrow \{Y\}$ for all $Y \in \mathrm{Adj}_{\mathrm{s}}(\mathcal{H}, X)$
foreach $X \in \mathcal{X}$ *and* $\mathcal{C} \in \widehat{\mathrm{Dep}}(X)$ **do**
  if $(\mathcal{C} \cup \{X\}) \setminus \{Y\} \in \widehat{\mathrm{Dep}}(Y)$ *and* $\widehat{\mathbf{Dep}^{\circ}}(\mathcal{C} \setminus \{Y\}, Y) = \mathcal{C} \setminus \{Y\}$ *for every* $Y \in \mathcal{C}$ **then**
    $\mathrm{Sym} \leftarrow \mathrm{Sym} \cup \{\mathcal{C} \cup \{X\}\}$
foreach *eligible* $(X, \mathcal{C}_1, \mathcal{C}_2)$ **do**
  if $\mathsf{Csep}_X(\mathcal{C}_1, \mathcal{C}_2)$ **then**
    Orient $V - X$ as $V \to X$ for all $V \in \mathcal{C}_1 \cup \mathcal{C}_2$
foreach $X \in \mathcal{X}$, $\mathcal{C} \in \widehat{\mathrm{Dep}}_{\geq 2}(X)$, *and* $Y \in \widehat{\mathbf{Dep}^{\circ}}(\mathcal{C}, X)$ **do**
  if X *and* Y *have no non-dashed edge* **then**
    Add the dashed edge $X -- Y$
while *a solid undirected edge is compelled by acyclicity or no-new-immorality* **do**
  Orient one such solid edge
return $\mathcal{H}$, $\widehat{\mathrm{Dep}}$, $\widehat{\mathbf{Dep}^{\circ}}$, and Sym