ASP-based Discovery of Semi-Markovian Causal Models under Weaker Assumptions

Zhalama[1*], **Jiji Zhang**[2], **Frederick Eberhardt**[3], **Wolfgang Mayer**[1] and **Mark Junjie Li**[4]
[1]University of South Australia
[2]Lingnan University
[3]California Institute of Technology
[4]Shenzhen University
zhalama@mymail.unisa.edu.au

Abstract

In recent years the possibility of relaxing the so-called Faithfulness assumption in automated causal discovery has been investigated. The investigation showed (1) that the Faithfulness assumption can be weakened in various ways that in an important sense preserve its power, and (2) that weakening of Faithfulness may help to speed up methods based on Answer Set Programming. However, this line of work has so far only considered the discovery of causal models without latent variables. In this paper, we study weakenings of Faithfulness for constraint-based discovery of semi-Markovian causal models, which accommodate the possibility of latent variables, and show that both (1) and (2) remain the case in this more realistic setting.

1 Introduction

Causal inference is of great interest in many scientific areas, and automated discovery of causal structure from data is drawing increasingly more attention in the field of machine learning. One of the standard approaches to automated causal discovery, known as the constraint-based approach, seeks to infer from data statistical relations among a set of random variables, and translate those relations into constraints on the underlying causal structure so that features of the causal structure may be determined from the constraints [Spirtes *et al.*, 2000; Pearl, 2000]. In this approach, the most commonly used constraints are in the form of conditional (in)dependence, which can serve as constraints on the causal structure due in the first place to the well known causal Markov assumption. The assumption states roughly that a causal structure, as represented by a directed acyclic graph (DAG), entails a certain set of conditional independence statements. With this assumption, a conditional dependency found in the data constrains the causal DAG.

The causal Markov assumption is almost universally accepted by researchers on causal discovery. However, by itself the assumption is too weak to enable interesting causal inference [Zhang, 2013]. It is therefore usually supplemented with an assumption known as Faithfulness, which states roughly that unless entailed by the causal structure according to the Markov assumption, no conditional independence relation should hold. With this assumption, conditional independence relations found in the data also constrain the causal DAG.

Unlike the causal Markov assumption, the Faithfulness assumption is often regarded as questionable. The standard defense of the assumption is that violations of Faithfulness involve fine-tuning of parameters (such as two causal pathways balancing out exactly), which is very unlikely if we assume parameter values are somehow randomly chosen. However, parameter values may not be randomly chosen, especially in situations where balancing of multiple causal pathways may be part of the design. More importantly, even if the true distribution is faithful to the true causal structure, with finite data, "apparent violations" of faithfulness can result from errors in statistical tests, when a false hypothesis of conditional independence fails to be rejected. Such apparent violations of faithfulness cannot be reasonably assumed away [Uhler *et al.*, 2013] and will bring troubles to causal discovery that assumes Faithfulness [Meek, 1996; Robins *et al.*, 2003].

For these reasons, in recent years the possibility of relaxing the Faithfulness assumption has been investigated [Ramsey *et al.*, 2006; Zhang and Spirtes, 2008; Zhang, 2013; Spirtes and Zhang, 2014; Raskutti and Uhler, 2018; Forster *et al.*, 2017]. This line of work made it clear that in the context of learning causal models with no latent variables, the Faithfulness assumption can be weakened or generalized in a number of ways while retaining its inferential power, because in theory these assumptions all reduce to the Faithfulness assumption when the latter happens to hold.

On a more practical note, causal discovery algorithms have also been developed to fit some of these weaker assumptions, most notably the Conservative PC algorithm [Ramsey *et al.*, 2006] and the greedy permutation-based algorithms [Wang *et al.*, 2017; Solus *et al.*, 2017]. More systematically, [Zhalama *et al.*, 2017] implemented and compared a number of weakenings of Faithfulness in the flexible approach to causal discovery based on Answer Set Programming (ASP) [Hyttinen *et al.*, 2014]. Among other things, they found, rather surprisingly, that some weakenings significantly boost the time efficiency of ASP-based algorithms. Since the main drawback of the ASP-based approach lies with its feasibility, this finding is potentially consequential for the further development

[*]Contact Author

of this approach.

However, neither the theoretical investigation nor the ASP-based practical exploration went beyond the limited (and unrealistic) context of learning causal models in the absence of latent confounding, also known as causal discovery with the assumption of *causal sufficiency* [Spirtes *et al.*, 2000]. Since latent confounding is ubiquitous, it is a serious limitation to restrict the study to causally sufficient settings. And it is especially unsatisfactory from the perspective of the ASP-based approach, which boasts the potential to deal with a most general search space that accommodates the possibility of latent confounding and that of causal loops [Hyttinen *et al.*, 2013].

In this paper, we make a step towards remedying this limitation by generalizing the aforementioned investigation in a setting where latent confounding is allowed (but not causal loops; we remark on a complication that will arise in the presence of causal loops in the end.) Since the investigation appeals to the ASP-based platform, we will follow previous work on this topic to use semi-Markovian causal models to represent causal structures with latent confounders. Among other things, we show that it remains the case that (1) the Faithfulness assumption can be weakened in various ways that in an important sense preserve its power, and (2) weakening of Faithfulness may help to speed up ASP-based methods.

The remainder of the paper will proceed as follows. In Section 2, we introduce terminologies and describe the basic setup. In Section 3, we review a few ways to relax the Faithfulness assumption that have been proposed in the context of causal discovery with causal sufficiency and have been proved to be *conservative* in a sense we will specify. Then, in Section 4, we discuss the complications that arise with semi-Markovian causal models, and establish generalizations of the results mentioned in Section 3. This is followed by a discussion in Section 5 of how to implement the weaker assumptions in the ASP platform. Finally, we report some simulation results in Section 6 that demonstrate the speed-up mentioned above, and conclude in Section 7.

2 Preliminaries

In this paper, the general graphical representation of a causal structure is by way of a *mixed graph*. The kind of mixed graph we will use is a triple $(\mathbf{V}, \mathbf{E}_1, \mathbf{E}_2)$, where $\mathbf{V}$ is a set of vertices (each representing a random variable), $\mathbf{E}_1$ a set of directed edges ($\rightarrow$) and $\mathbf{E}_2$ a set of bi-directed edges ($\leftrightarrow$). In general, more than one edge is allowed between two vertices, but no edge is allowed between a vertex and itself. Two vertices are said to be *adjacent* if there is at least one edge between them. Given an edge $X \rightarrow Y$, X is called a *parent* of Y and Y a *child* of X. We also say the edge has a *tail* at X and an *arrowhead* at Y. An edge $X \leftrightarrow Y$ is said to have an arrowhead at both X and Y. A *path* between X and Y consists of an ordered sequence of distinct vertices $\langle X = V_1, ..., V_n = Y \rangle$ and a sequence of edges $\langle E_1, ..., E_{n-1} \rangle$ such that for $1 \leq i \leq n-1$, E_i is an edge between V_i and V_{i+1}. Such a path is a *directed path from X to Y* if for all $1 \leq i \leq n-1$, E_i is a directed edge from V_i to V_{i+1}. X is an *ancestor* of Y and Y an *descendant* of X, if either $X = Y$ or there is a directed path from X to Y. A *directed cycle* occurs when two distinct vertices are ancestors of each other.

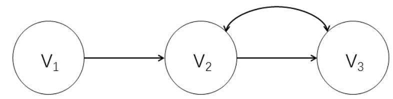

Figure 1: An inducing path between two non-adjacent vertices

If a mixed graph does not contain any directed cycle, we will call it a *semi-Markovian causal model* (SMCM), also known as an *acyclic directed mixed graph* (ADMG). Intuitively a directed edge in an SMCM represents a direct causal relationship, and a bi-directed edge represents the presence of latent confounding. A directed acyclic graph (DAG) is a special case where no bi-directed edge appears. A DAG can be thought of as representing a causal model over a *causally sufficient* set of random variables, which may be referred to as a Markovian causal model (MCM).

The conditional independence statements entailed by a graph can be determined graphically by a separation criterion. One statement of this criterion is *m-separation*, which is a natural generalization of the celebrated *d-separation* criterion for DAGs [Pearl, 1988]. Given any path in a mixed graph G, a non-endpoint vertex V on the path is said to be a *collider* on the path if both edges incident to V on the path have an arrowhead at V. Otherwise it is said to be a *non-collider* on the path.

Definition 1 (m-connection and m-separation)**.** Given a mixed graph G over $\mathbf{V}$ and $\mathbf{Z} \subseteq \mathbf{V}$, a path in G is *m-connecting given* $\mathbf{Z}$ if every non-collider on the path is not in $\mathbf{Z}$ and every collider on the path has a descendant in $\mathbf{Z}$.

For any distinct $X, Y \notin \mathbf{Z}$, X and Y are *m-separated by* $\mathbf{Z}$ in G (written as $X \perp_G Y \mid \mathbf{Z}$) if there is no path between X and Y that is m-connecting given $\mathbf{Z}$. Otherwise X and Y are said to be *m-connected by* $\mathbf{Z}$.

For any $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ that are pairwise disjoint, $\mathbf{X}$ and $\mathbf{Y}$ are m-separated by $\mathbf{Z}$ in G if every vertex in $\mathbf{X}$ and every vertex in $\mathbf{Y}$ are m-separated by $\mathbf{Z}$.

This definition obviously reduces to that of d-connection and d-separation in the case of DAGs. It is well known that in a DAG, two vertices are adjacent if and only if no set of other vertices d-separates them. The 'only if' direction holds for SMCMs, but the 'if' direction does not. For example, in the simple SMCM in Figure 1, V_1 and V_3 are not adjacent, but neither the empty set nor the set $\{V_2\}$ m-separates them. This motivates the following definition.

Definition 2 (inducing path)**.** A path between X and Y is an *inducing path* if every non-endpoint vertex on the path is a collider and also an ancestor of either X or Y.

For example, in Figure 1, the path $V_1 \rightarrow V_2 \leftrightarrow V_3$ is an inducing path between V_1 and V_3. In general, two vertices in an SMCM are not m-separated by any set of other variables if and only if there is an inducing path between them [Verma, 1993]. Note that an edge between two vertices constitutes an inducing path. Following [Richardson, 1997], we call two vertices *virtually adjacent* if there is an inducing path between them. Adjacency entails virtual adjacency, but not vice versa.

3 Faithfulness and Its Weakening for Learning Causal Models without Latent Variables

We now review some proposals of weakening the Faithfulness assumption in the context of learning (acyclic) causal structures in the absence of latent confounding. In such a case, the target is a DAG over the given set of random variables $\mathbf{V}$, in which each edge represents a direct causal relation relative to $\mathbf{V}$ [Spirtes *et al.*, 2000]. Let G denote the unknown true causal DAG over $\mathbf{V}$, and P denote the true joint probability distribution over $\mathbf{V}$. The causal Markov assumption can be formulated as:

Causal Markov assumption For every pairwise disjoint $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, if $\mathbf{X} \perp_G \mathbf{Y} \mid \mathbf{Z}$, then $\mathbf{X} \perp\!\!\!\perp_P \mathbf{Y} \mid \mathbf{Z}$.

where '$\mathbf{X} \perp_G \mathbf{Y} \mid \mathbf{Z}$' means that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in G, and '$\mathbf{X} \perp\!\!\!\perp_P \mathbf{Y} \mid \mathbf{Z}$' means that $\mathbf{X}$ and $\mathbf{Y}$ are independent conditional on $\mathbf{Z}$ according to P.

The converse is the causal Faithfulness assumption:

Causal Faithfulness assumption For every pairwise disjoint $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, if $\mathbf{X} \perp\!\!\!\perp_P \mathbf{Y} \mid \mathbf{Z}$, then $\mathbf{X} \perp_G \mathbf{Y} \mid \mathbf{Z}$.

As mentioned earlier, the Faithfulness assumption is regarded as much more questionable than the Markov assumption, and the literature has seen a number of proposals to relax it. In this paper, we focus on the following three.[1]

Adjacency-faithfulness assumption For every distinct $X, Y \in \mathbf{V}$, if X and Y are adjacent in G, then $X \not\perp\!\!\!\perp_P Y \mid \mathbf{Z}$, for every $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$.

Number-of-Edges(NoE)-minimality assumption: G is NoE-minimal in the sense that no DAG with a smaller number of edges than G satisfies the Markov assumption with P.

Number-of-Independencies(NoI)-minimality assumption: G is NoI-minimal in the sense that no DAG that entails a greater number of conditional independence statements than G does, satisfies the Markov assumption with P.

Under the Markov assumption, these assumptions are all weaker than the Faithfulness assumption. In words, Adjacency-faithfulness says that two variables that are adjacent in the causal structure are dependent given any conditioning set. It was first introduced in [Ramsey *et al.*, 2006] and motivated the CPC (conservative PC) algorithm. NoE-minimality says that the true causal structure has the least number of edges among all structures that satisfy the Markov assumption. It underlies the novel permutation-based algorithms that were developed recently [Raskutti and Uhler, 2018; Wang *et al.*, 2017; Solus *et al.*, 2017]. NoI-minimality says that the true causal structure entails the greatest number of conditional independence statements among all structures that satisfy the Markov assumption. In the ASP-based methods, the 'hard-deps' conflict resolution scheme in [Hyttinen *et al.*, 2014] happened to implement this minimality constraint.

Theoretically these assumptions are particularly interesting because although they are weaker than Faithfulness (given the Markov assumption), they are in a sense strong enough to preserve the inferential power of Faithfulness. It has been shown that when Faithfulness happens to hold, all these weaker assumptions become equivalent to Faithfulness [Zhalama *et al.*, 2017]. In other words, while they are weaker than Faithfulness and therefore still hold in many cases when Faithfulness does not, they rule out exactly the same causal graphs as Faithfulness does when the latter happens to be satisfied. We propose to call this kind of weakening *conservative*, for it retains the inferential power of Faithfulness whenever Faithfulness is applicable. The choice between a stronger assumption and a weaker one usually involves a trade-off between risk (of making a false assumption) and inferential power, but there is no such trade-off if the weakening is conservative.

In addition to this theoretical virtue, both Adjacency-faithfulness and NoE-minimality, and especially Adjacency-faithfulness, have been shown to significantly improve the time efficiency of ASP-based causal discovery methods, without significant sacrifice in performance [Zhalama *et al.*, 2017]. We aim to extend these findings to the much more realistic setting where latent confounding may be present.

4 Weakening Faithfulness for Learning Semi-Markovian Causal Models

When the set of observed variables $\mathbf{V}$ is not causally sufficient, which means that some variables in $\mathbf{V}$ share a common cause or confounder that is not observed, it is no longer appropriate to represent the causal structure in question with a DAG over $\mathbf{V}$. One option is to explicitly invoke latent variables in the representation and assume the underlying causal structure is properly represented by a DAG over $\mathbf{V}$ plus some latent variables $\mathbf{L}$. Another option is to suppress latent variables and use bi-directed edges to represent latent confounding. The use of SMCMs exemplifies the latter approach.[2]

As [Verma, 1993] showed, for every DAG over $\mathbf{V}$ and set of latent variables $\mathbf{L}$, there is a unique projection into an SMCM over $\mathbf{V}$ that preserves both the causal relations among $\mathbf{V}$ and the entailed conditional independence relations among $\mathbf{V}$. Moreover, as [Richardson, 2003] pointed out, the original causal DAG with latent variables and its projection into an SMCM are equivalent regarding the (nonparametric) identification of causal effects. These facts justify using SMCMs to represent causal structures with latent confounding.

So let us suppose the underlying causal structure over $\mathbf{V}$ is properly represented by an SMCM G and let P denote

[1] Another two proposals are known as 'Triangle-Faithfulness plus SGS-minimality' [Spirtes and Zhang, 2014] and 'P-minimality' [Zhang, 2013]. It is not yet clear how to implement the latter in ASP, and the former did not seem to help much with ASP-based methods [Zhalama *et al.*, 2017].

[2] Another important example is the use of ancestral graph Markov models [Richardson and Spirtes, 2002].

the true joint distribution over $\mathbf{V}$. In this setting, the causal Markov and Faithfulness assumptions can be formulated as before (in Section 3), except that the separation criterion is now understood as the more general m-separation. Next we examine the proposals of weakening Faithfulness.

Regarding Adjacency-faithfulness, it is easy to see that it remains a logical consequence of Faithfulness. If two variables are adjacent in an SMCM, then given any set of other variables, the two are m-connected (any edge between them constitutes a m-connecting path). Thus, if Faithfulness holds, then they are not independent conditional on any set of other variables, exactly what is required by Adjacency-faithfulness. Since Adjacency-faithfulness does not entail Faithfulness in the case of DAGs and DAGs are special cases of SMCMs, Adjacency-faithfulness remains weaker than Faithfulness.

However, it is now too weak to be a conservative weakening of Faithfulness. Here is a very simple example. Suppose the true causal structure over three random variables is a simple causal chain $V_1 \to V_2 \to V_3$, and suppose the joint distribution is Markov and Faithful to this structure. So we have $V_1 \perp\!\!\!\perp V_3 \mid V_2$. Then the distribution is not Faithful to the structure in Figure 1, because that structure does not entail that V_1 and V_3 are m-separated by V_2. Still, Adjacency-faithfulness is satisfied by the distribution and the structure in Figure 1, for the only violation of Faithfulness occurs with regard to V_1 and V_3, which are not adjacent. Therefore, in this simple case where Faithfulness happens to hold, if we just assume Adjacency-faithfulness, we are not going to rule out the structure in Figure 1, which would be ruled out if we assumed Faithfulness.

This simple example suggests that we should consider the following variation:

V(irtual)-adjacency-faithfulness assumption: For every distinct $X, Y \in \mathbf{V}$, if X and Y are virtually adjacent in G (i.e., if there is an inducing path between X and Y in G), then $X \not\perp\!\!\!\perp_P Y \mid \mathbf{Z}$, for every $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$.

V-adjacency-faithfulness is obviously stronger than Adjacency-faithfulness, but we can prove that it remains weaker than Faithfulness. More importantly, it is strong enough to be a conservative weakening of Faithfulness.

How about NoE-minimality? Since more than one edge can appear between two vertices, NoE-minimality (as it is formulated in Section 3) is no longer a consequence of Faithfulness. To see this, just suppose the true structure over two random variables is simply $V_1 \to V_2$ together with $V_1 \leftrightarrow V_2$ (i.e., V_1 is a cause of V_2 but the relation is also confounded), and suppose the distribution is Markov and Faithful to this structure. NoE-minimality is violated here, for taking away either (but not both) of the two edges still results in a structure that satisfies the Markov assumption.

So NoE-minimality is not a weakening of Faithfulness. Note that in the case of DAGs, minimization of the number of edges is equivalent to minimization of the number of adjacencies. If we replace the former with the latter, the above example is taken care of (for taking away the adjacency in that example will result in a structure that fails the Markov assumption). However, it is also easy to construct an example where an adjacency in an SMCM can be taken away without affecting the independence model [Richardson and Spirtes, 2002], so adjacency-minimality also fails to be a weakening of Faithfulness. The right generalization of NoE-minimality is unsurprisingly the following:

V(irtual)-adjacency-minimality assumption: G is V-adjacency-minimal in the sense that no SMCM with a smaller number of virtual adjacencies than G satisfies the Markov assumption with P.

Finally, Since NoI-minimality is concerned with entailed conditional independence statements, it is straightforwardly generalized to the setting of SMCMs (just replace 'DAG' with 'SMCM' in the original formulation), and remains a conservative weakening of Faithfulness. Here then is the main result of this section.[3]

Theorem. Given the causal Markov assumption, the V-adjacency-faithfulness assumption, V-adjacency-minimality assumption, and NoI-minimality assumptions are all conservative weakenings of the Faithfulness assumption, in the following sense: for each of the 3 assumptions **AS**,

(a) **AS** is entailed by, but does not entail, Faithfulness.

(b) For every joint probability distribution P over $\mathbf{V}$, if there exists an SMCM that satisfies both Markov and Faithfulness assumptions with P, then for every SMCM G that satisfies the Markov assumption with P, G satisfies Faithfulness if and only if G satisfies **AS** with P.

5 ASP-based Causal Discovery of SMCMs

We instantiated causal discovery algorithms, which adopt V-adjacency-faithfulness and V-adjacency-minimality, using the framework of [Hyttinen *et al.*, 2014]. This framework offers a generic constraint-based causal discovery method based on Answer Set Programming (ASP). The logic is used to define Boolean atoms that represent the presence or absence of a directed or bi-directed edge in an SMCM. In addition, conditional independence/dependence statements (CI/CDs) obtained from tests on the input data are encoded in this logic. Finally, background assumptions, such as Markov and Faithfulness, are written as logical constraints enforcing a correspondence between the encoded test results and the underlying Boolean atoms (the edges of the SMCM). Solutions, which are truth-value assignments to the Boolean atoms, satisfying such a correspondence are found using off-the-shelf solvers. The set of solutions specifies the set of SMCMs that satisfy all the input CI/CDs and the background assumptions. Given that the results of the statistical tests may conflict with the background assumptions, there may be no solution, i.e. there is no SMCM that satisfies all the input CI/CDs and background assumptions. For that case [Hyttinen *et al.*, 2014] introduced the following optimization to resolve the conflict:

$$G^* \in \arg\min_{G \in \mathcal{G}} \sum_{k \in \mathbf{K} \text{ s.t. } G \not\models k} w(k) \quad (1)$$

[3] For a proof of the theorem, we refer readers to the appendices in the expanded version of the paper [Zhalama *et al.*, 2019].

In words, an output graph G^* minimizes the weighted sum of input CI/CDs, which it does not satisfy given the encoded background assumptions. [Hyttinen *et al.*, 2014] adopted three weighting schemes for the weights $w(.)$: (1) "constant weights" (CW) assigns a weight of 1 to each CI and CD constraint. (2) "hard dependencies weights" (HW/NoI-m) assigns infinite weight to any observed CD, and a weight of 1 to any CI. (3) "log weights" (LW) is a pseudo-Bayesian weighting scheme, where the weights depend on the log posterior probability of the CI/CDs being true (see their Sec. 4).

To encode V-adjacency-faithfulness and V-adjacency-minimality, we need to encode in ASP what it is for an SMCM to have an inducing path and a virtual adjacency, and then replace the encoding of the Faithfulness assumption in [Hyttinen *et al.*, 2014] with its weaker versions. Figure 2 summarizes the ASP-encoding of V-adjacency-faithfulness and V-adjacency-minimality. We briefly explain the predicates:

- $edge(X, Z)$ and $conf(X, Z)$: $X \rightarrow Z$ and $X \leftrightarrow Z$, respectively, are in the SMCM.
- $ancestors(Z, X, Y)$: Z is an ancestor of X or Y in the SMCM.
- $h(X, Z, Y)$: There is a path between X and Z which is into Z, and if the path consists of two or more edges, every non-endpoint vertex on the path is a collider and every vertex on the path is an ancestor of either X or Y.
- $t(X, Z, Y)$: It differs from $h(X, Z, Y)$ only in that the path between X and Z is out of Z. Together, $t(.)$ and $h(.)$ are used to specify the possible inducing paths.
- $vadj(X, Y)$: X and Y are virtually adjacent.

Inference rules for virtual-adjacency:

$$h(X, Z, Y) \text{ :- } edge(X, Z), ancestors(Z, X, Y).$$
$$h(X, Z, Y) \text{ :- } conf(X, Z), ancestors(Z, X, Y).$$
$$h(X, Z, Y) \text{ :- } h(X, U, Y), conf(Z, U), ancestors(Z, X, Y).$$
$$t(X, Z, Y) \text{ :- } h(X, U, Y), edge(Z, U), ancestors(Z, X, Y).$$
$$vadj(X, Y) \text{ :- } h(X, Y, Y).$$
$$vadj(X, Y) \text{ :- } t(X, Y, Y).$$
$$vadj(X, Y) \text{ :- } edge(Y, X).$$

Virtual-adjacency-faithfulness (violations):

$\forall X \forall Y > X, \forall \mathbf{C} \subseteq \mathbf{V} \setminus \{X, Y\}$,
:- not $vadj(X, Y), indep(X, Y, \mathbf{C}, w)$

Virtual-adjacency-minimality (optimization of weak constraints):

$\forall X \forall Y > X$,
$$fail(X, Y, w = 1) \text{ :- } vadj(X, Y).$$
$$:\sim fail(X, Y, w).\ [w]$$

(Variables are in an arbitrary order so that $indep(X, Y, \mathbf{C}, w)$ and $dep(X, Y, \mathbf{C}, w)$ are considered only if $Y > X$, in order to avoid double counting.)

Figure 2: ASP Encoding of V-adjacency-faithfulness and V-adjacency-minimality

- $indep(X, Y, \mathbf{C}, w)$: X and Y are independent conditional on C, given as input fact, with weight w.

For V-adjacency-faithfulness, we encode that any CI statement $X \perp\!\!\!\perp Y \mid \mathbf{C}$ implies that X and Y are not virtually adjacent. For V-adjacency-minimality, we employ the minimization of the number of virtual-adjacencies. By encoding the weaker assumptions in the framework of [Hyttinen *et al.*, 2014], we then have the following algorithms (Hyttinen et al.'s algorithm based on the 'hard dependencies' weights is equivalent to one based on NoI-minimality):

- **VadjF**: Virtual-adjacency-faithfulness + Markov
- **VadjM**: Virtual-adjacency-minimality + Markov

6 Simulations

We report two types of simulation, one using an independence oracle that specifies the true CI/CDs of the causal model, and one that uses the CI/CDs inferred from the sample data.

For both simulations we followed the model generation process of [Hyttinen *et al.*, 2014] for causally insufficient models: We generated 100 random linear Gaussian models over 6 vertices with an average edge degree of 1 for directed edges. The edge coefficients were drawn uniformly from $[-0.8, -0.2] \cup [0.2, 0.8]$. The error covariance matrices (which also represent the confounding) were generated using the observational covariance matrix of a similarly constructed causally sufficient model (with its error covariances sampled from $N(0.5, 0.01)$).

In the oracle setting, we randomly generated 100 linear Gaussian models with latent confounders over 6 variables and then input the independence oracles implied by these models. We observed that the algorithms based on V-adjacency-faithfulness, on V-adjacency-minimality, and on NoI-minimality (which is equivalent to using 'hard dependencies' weighting) all returned the exact same results as the algorithm based on Faithfulness did, which is consistent with the theoretical results in Section 4 and confirms the correctness of our encoding.

In the finite sample case we generated five data sets with 500 samples from each of the 100 models. We used correlational t-tests and tried 10 threshold values for rejecting the null hypothesis (0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25). The test results formed the input for the algorithms. We also used the log-weighting scheme and tried 10 values for the free parameter of the Bayesian test $(0.05, 0.09, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9)$.

For each algorithm we output all possible solutions and compared the d-connections common to all the output graphs against those of the true data generating graph. In all the 100(models)$*$5(data)$*$10(parameters) $= 5,000$ runs, Faithfulness was satisfied in only $367/5000$ of the cases while V-adjacency-faithfulness was satisfied in $2065/5000$ of the cases. This shows that V-adjacency-faithfulness is indeed significantly weaker than Faithfulness and greatly reduces the number of conflicts. By definition, V-adjacency-minimality can always be satisfied. Figure 3 plots the ROC curves for the inferred d-connections. Under "constant weighting" (CW)

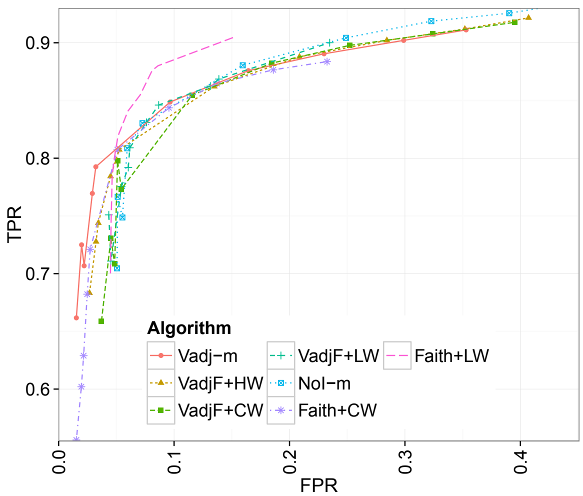

Figure 3: ROC of d-connections

"hard dependencies weighting" (HW), using V-adjacency-faithfulness achieves comparable accuracy to using faithfulness, with some trade-offs between false-positive rates and true-positive rates. Under "log weighting" (LW), however, using Faithfulness seems slightly more accurate than using V-adjacency-faithfulness, though using V-adjacency-minimality seems to generally yield the lowest false-positive rates. How to adapt the "log weighting" to fit V-adjacency-faithfulness better is an interesting question for future work.

Finally, to explore the efficiency gain of the weakened faithfulness assumptions, we generated 100 random linear Gaussian models with latent confounders over 8 variables and generated one data set with 500 samples from each model. For each algorithm, we only required that one graph be found. Figure 4 shows the sorted solving times for the different background assumptions (with maximum time budget of 5,000s). As in the causally sufficient case, we see a significant improvement in solving times when using the weakened faithfulness assumptions.

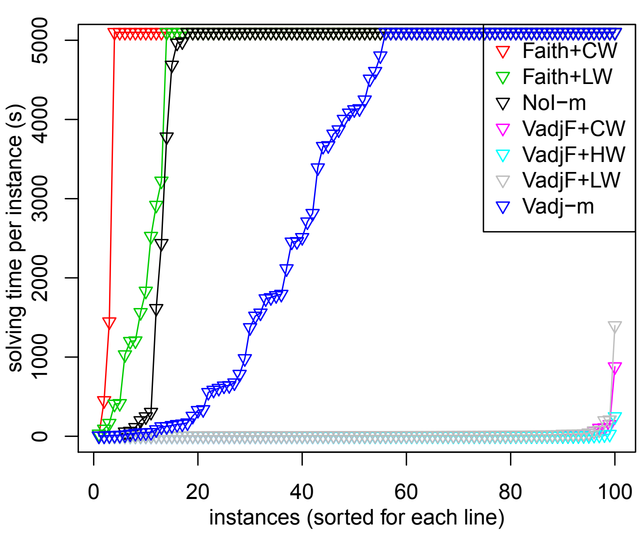

Figure 4: Sorted Solving Times for 8 Variables (time-out at 5,000s)

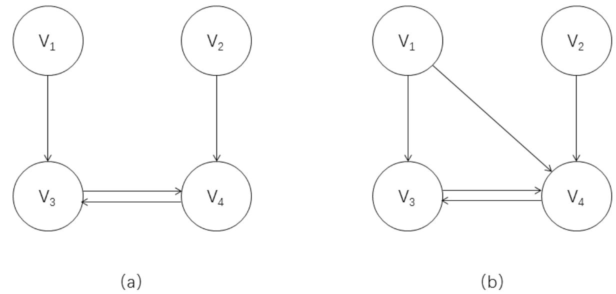

Figure 5: Illustration of a complication in cyclic models

7 Conclusion

We have shown how to extend the results on weakening Faithfulness in the context of learning causal DAGs to the more realistic context of learning SMCMs that allow for the representation of unmeasured confounding. We identified generalizations of some proposals of weakening Faithfulness in the literature and showed that they continue to be what we call conservative weakenings. Moreover, we implemented ASP-based algorithms for learning SMCMs based on these weaker assumptions. The simulation results suggest that some of these weaker assumptions, especially V-adjacency-faithfulness, help to save solving time in ASP-based algorithms to a significant extent.

In this connection, a direction of future work is to explore how the apparent advantage of using weaker assumptions may be realized on top of other ASP-based causal discovery methods, such as ETIO in [Borboudakis and Tsamardinos, 2016] and ACI in [Magliacane *et al.*, 2016].

One great appeal of the ASP-based approach is that the background assumptions that determine the search space can be flexibly adjusted to include causal models with both latent confounding and causal feedback. We close with an illustration of a (further) complication that arises in cyclic causal models. Suppose the true causal structure is the cyclic one in Figure 5(a), which entails that $V_1 \perp V_2$ and $V_1 \perp V_2 \mid \{V_3, V_4\}$. Suppose the true distribution is Markov and Faithful to this structure and hence features exactly two non-trivial conditional independencies. Then the distribution is not Faithful to the structure in Figure 5(b) (for that structure does not entail $V_1 \perp V_2 \mid \{V_3, V_4\}$), but it is still V-adjacency-faithful (for V_1 and V_2 are not virtually adjacent).

This means that even V-adjacency-faithfulness is not a conservative weakening of Faithfulness when causal feedback is allowed. Whether it can be strengthened into a useful conservative weakening for the purpose of learning cyclic models is worth further investigation.

Acknowledgments

Jiji Zhang was supported by GRF LU13602818 from the RGC of Hong Kong. Frederick Eberhardt was supported by NSF grant 1564330.

References

[Borboudakis and Tsamardinos, 2016] Giorgos Borboudakis and Ioannis Tsamardinos. Towards robust and versatile causal discovery for business applications. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1435–1444. ACM, August 2016.

[Forster *et al.*, 2017] Malcolm Forster, Garvesh Raskutti, Reuben Stern, and Naftali Weinberger. The frugal inference of causal relations. *British Journal for the Philosophy of Science*, 69(3):821–848, April 2017.

[Hyttinen *et al.*, 2013] Antti Hyttinen, Patrik Hoyer, Frederick Eberhardt, and Matti Järvisalo. Discovering cyclic causal models with latent variables: A general SAT-based procedure. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 301–310. AUAI Press, August 2013.

[Hyttinen *et al.*, 2014] Antti Hyttinen, Frederick Eberhardt, and Matti Järvisalo. Constraint-based causal discovery: Conflict resolution with Answer Set Programming. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 340–349. AUAI Press, July 2014.

[Magliacane *et al.*, 2016] Sara Magliacane, Tom Claassen, and Joris Mooij. Ancestral causal inference. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 4466–4474. Curran Associates Inc., December 2016.

[Meek, 1996] Christopher Meek. *Graphical Causal Models: Selecting Causal and Statistical Models*. PhD thesis, Department of Philosophy, Carnegie Mellon University, 1996.

[Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.

[Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Oxford University Press, 2000.

[Ramsey *et al.*, 2006] Joseph Ramsey, Peter Spirtes, and Jiji Zhang. Adjacency-faithfulness and conservative causal inference. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 401–408. AUAI Press, July 2006.

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

[Richardson and Spirtes, 2002] Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

[Richardson, 1997] Thomas Richardson. A characterization of Markov equivalence for directed cyclic graphs. *International Journal of Approximate Reasoning*, 17(2-3):107 – 162, August–October 1997.

[Richardson, 2003] Thomas Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, March 2003.

[Robins *et al.*, 2003] James Robins, Richard Scheines, Peter Spirtes, and Larry Wasserman. Uniform consistency in causal inference. *Biometrika*, 90:491–515, September 2003.

[Solus *et al.*, 2017] Liam Solus, Yuhao Wang, Lenka Matejovicova, and Caroline Uhler. Consistency guarantees for permutation-based causal inference algorithms. *arXiv:1702.03530*, July 2017.

[Spirtes and Zhang, 2014] Peter Spirtes and Jiji Zhang. A uniformly consistent estimator of causal effects under the k-triangle-faithfulness assumption. *Statistical Science*, 29(4):662–678, November 2014.

[Spirtes *et al.*, 2000] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction and Search*. MIT Press, 2 edition, 2000.

[Uhler *et al.*, 2013] Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of faithfulness assumption in causal inference. *The Annals of Statistics*, 41:436–463, April 2013.

[Verma, 1993] Thomas Verma. Graphical aspects of causal model. Technical report, Computer Science Department, University of California, Los Angles, 1993.

[Wang *et al.*, 2017] Yuhao Wang, Liam Solus, Karren Dai Yang, and Caroline Uhler. Permutation-based causal inference algorithms with interventions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5824–5833. Curran Associates Inc., December 2017.

[Zhalama *et al.*, 2017] Zhalama, Jiji Zhang, Frederick Eberhardt, and Wolfgang Mayer. Sat-based causal discovery under weaker assumptions. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence*, pages 630–639. AUAI Press, June 2017.

[Zhalama *et al.*, 2019] Zhalama, Jiji Zhang, Frederick Eberhardt, Wolfgang Mayer, and Mark Junjie Li. ASP-based discovery of semi-Markovian causal models under weaker assumptions. *arXiv:1906.02385*, June 2019.

[Zhang and Spirtes, 2008] Jiji Zhang and Peter Spirtes. Detection of unfaithfulness and robust causal inference. *Minds and Machines*, 18(2):239–271, June 2008.

[Zhang, 2013] Jiji Zhang. A comparison of three Occam's razor for Markovian causal models. *British Jounrnal for the Philosophy of Science*, 64(2):423–448, June 2013.