# Learning AMP Chain Graphs under Faithfulness

Jose M. Peña

ADIT, IDA, Linköping University, SE-58183 Linköping, Sweden

jose.m.pena@liu.se

## Abstract

This paper deals with chain graphs under the alternative Andersson-Madigan-Perlman (AMP) interpretation. In particular, we present a constraint based algorithm for learning an AMP chain graph a given probability distribution is faithful to. We also show that the extension of Meek's conjecture to AMP chain graphs does not hold, which compromises the development of efficient and correct score+search learning algorithms under assumptions weaker than faithfulness.

## 1 Introduction

This paper deals with chain graphs (CGs) under the alternative Andersson-Madigan-Perlman (AMP) interpretation (Andersson et al., 2001). In particular, we present an algorithm for learning an AMP CG a given probability distribution is faithful to. To our knowledge, we are the first to present such an algorithm. However, it is worth mentioning that, under the classical Lauritzen-Wermuth-Frydenberg (LWF) interpretation of CGs (Lauritzen, 1996), such an algorithm already exists (Ma et al., 2008; Studený, 1997). Moreover, we have recently developed an algorithm for learning LWF CGs under the milder composition property assumption (Peña et al., 2012).

The AMP and LWF interpretations of CGs are sometimes considered as competing and, thus, their relative merits have been pointed out (Andersson et al., 2001; Drton and Eichler, 2006; Levitz et al., 2001; Roverato and Studený, 2006). Note, however, that no interpretation subsumes the other: There are many independence models that can be induced by a CG under one interpretation but that cannot be induced by any CG under the other interpretation (Andersson et al., 2001, Theorem 6).

The rest of the paper is organized as follows. Section 2 reviews some concepts. Section 3 presents the algorithm. Section 4 proves its correctness. Section 5 closes with some discussion.

## 2 Preliminaries

In this section, we review some concepts from probabilistic graphical models that are used later in this paper. All the graphs and probability distributions in this paper are defined over a finite set $V$. All the graphs in this paper are hybrid graphs, i.e. they have (possibly) both directed and undirected edges. The elements of $V$ are not distinguished from singletons. We denote by $|X|$ the cardinality of $X \subseteq V$.

If a graph $G$ contains an undirected (resp. directed) edge between two nodes $V_1$ and $V_2$, then we write that $V_1 - V_2$ (resp. $V_1 \rightarrow V_2$) is in $G$. The parents of a set of nodes $X$ of $G$ is the set $pa_G(X) = \{V_1 | V_1 \rightarrow V_2 \text{ is in } G, V_1 \notin X \text{ and } V_2 \in X\}$. The neighbors of a set of nodes $X$ of $G$ is the set $ne_G(X) = \{V_1 | V_1 - V_2 \text{ is in } G, V_1 \notin X \text{ and } V_2 \in X\}$. The adjacents of a set of nodes $X$ of $G$ is the set $ad_G(X) = \{V_1 | V_1 \rightarrow V_2, V_1 - V_2 \text{ or } V_1 \leftarrow V_2 \text{ is in } G, V_1 \notin X \text{ and } V_2 \in X\}$. A route from a node $V_1$ to a node $V_n$ in $G$ is a sequence of (not necessarily distinct) nodes $V_1, \ldots, V_n$ such that $V_i \in ad_G(V_{i+1})$ for all $1 \leq i < n$. The length of a route is the number of (not necessarily distinct) edges in the route, e.g. the length of the route $V_1, \ldots, V_n$ is $n - 1$. A route is called a cycle if $V_n = V_1$. A route is called descending if $V_i \in pa_G(V_{i+1}) \cup ne_G(V_{i+1})$ for all $1 \leq i < n$. The descendants of a set of nodes $X$ of $G$ is the set $de_G(X) = \{V_n | \text{ there is a descending route from } V_1 \text{ to } V_n \text{ in } G, V_1 \in X \text{ and } V_n \notin X\}$. A cycle

is called a semidirected cycle if it is descending and $V_i \to V_{i+1}$ is in $G$ for some $1 \leq i < n$. A chain graph (CG) is a hybrid graph with no semidirected cycles. A set of nodes of a CG is connected if there exists a route in the CG between every pair of nodes in the set st all the edges in the route are undirected. A connectivity component of a CG is a connected set that is maximal wrt set inclusion. The connectivity component a node $A$ of a CG $G$ belongs to is denoted as $co_G(A)$. The subgraph of $G$ induced by a set of its nodes $X$ is the graph over $X$ that has all and only the edges in $G$ whose both ends are in $X$. An immorality is an induced subgraph of the form $A \to B \leftarrow C$. A flag is an induced subgraph of the form $A \to B - C$. If $G$ has an induced subgraph of the form $A \to B \leftarrow C$ or $A \to B - C$, then we say that the triplex $(\{A, C\}, B)$ is in $G$. Two CGs are triplex equivalent iff they have the same adjacencies and the same triplexes.

A node $B$ in a route $\rho$ is called a head-no-tail node in $\rho$ if $A \to B \leftarrow C$, $A \to B - C$, or $A - B \leftarrow C$ is a subroute of $\rho$ (note that maybe $A = C$ in the first case). Let $X$, $Y$ and $Z$ denote three disjoint subsets of $V$. A route $\rho$ in a CG $G$ is said to be $Z$-open when (i) every head-no-tail node in $\rho$ is in $Z$, and (ii) every other node in $\rho$ is not in $Z$. When there is no route in $G$ between a node in $X$ and a node in $Y$ that is $Z$-open, we say that $X$ is separated from $Y$ given $Z$ in $G$ and denote it as $X \perp_G Y | Z$.[1] We denote by $X \not\perp_G Y | Z$ that $X \perp_G Y | Z$ does not hold. Likewise, we denote by $X \perp_p Y | Z$ (resp. $X \not\perp_p Y | Z$) that $X$ is independent (resp. dependent) of $Y$ given $Z$ in a probability distribution $p$. The independence model induced by $G$, denoted as $I(G)$, is the set of separation statements $X \perp_G Y | Z$. We say that $p$ is Markovian with respect to $G$ when $X \perp_p Y | Z$ if $X \perp_G Y | Z$ for all $X$, $Y$ and $Z$ disjoint subsets of $V$. We say that $p$ is faithful to $G$ when $X \perp_p Y | Z$ iff $X \perp_G Y | Z$ for all $X$, $Y$ and $Z$ disjoint subsets of $V$. If two CGs $G$ and $H$ are triplex equivalent, then

$I(G) = I(H)$.[2]

Let $X$, $Y$, $Z$ and $W$ denote four disjoint subsets of $V$. Any probability distribution $p$ satisfies the following properties: Symmetry $X \perp_p Y | Z \Rightarrow Y \perp_p X | Z$, decomposition $X \perp_p Y \cup W | Z \Rightarrow X \perp_p Y | Z$, weak union $X \perp_p Y \cup W | Z \Rightarrow X \perp_p Y | Z \cup W$, and contraction $X \perp_p Y | Z \cup W \wedge X \perp_p W | Z \Rightarrow X \perp_p Y \cup W | Z$. Moreover, if $p$ is faithful to a CG, then it also satisfies the following properties: Intersection $X \perp_p Y | Z \cup W \wedge X \perp_p W | Z \cup Y \Rightarrow X \perp_p Y \cup W | Z$, and composition $X \perp_p Y | Z \wedge X \perp_p W | Z \Rightarrow X \perp_p Y \cup W | Z$.[3]

## 3   The Algorithm

Our algorithm, which can be seen in Table 1, resembles the well-known PC algorithm (Meek, 1995; Spirtes et al., 1993). It consists of two phases: The first phase (lines 1-8) aims at learning adjacencies, whereas the second phase (lines 9-10) aims at directing some of the adjacencies learnt. Specifically, the first phase declares that two nodes are adjacent iff they are not separated by any set of nodes. Note that the algorithm does not test every possible separator (see line 5). Note also that the separators tested are tested in increasing order of size (see lines 2, 5 and 8). The second phase consists of two steps. In the first step, the ends of some of the edges learnt in the first phase are blocked according to the rules R1-R4 in Table 2. A block is represented by a perpendicular line such as in $\vdash$ or $\rightarrowtail$, and it means that the edge cannot be directed in that direction. In the second step, the edges with exactly one unblocked end get directed in the direction of the unblocked end. The rules R1-R4 work as follows: If the conditions in the antecedent of a rule are satisfied, then the modifications in the consequent of the

---

[1]See (Andersson et al., 2001, Remark 3.1) for the equivalence of this and the standard definition of separation.

[2]To see it, note that there are Gaussian distributions $p$ and $q$ that are faithful to $G$ and $H$, respectively (Levitz et al., 2001, Theorem 6.1). Moreover, $p$ and $q$ are Markovian wrt $H$ and $G$, respectively, by Andersson et al. (2001, Theorem 5) and Levitz et al. (2001, Theorem 4.1).

[3]To see it, note that there is a Gaussian distribution that is faithful to $G$ (Levitz et al., 2001, Theorem 6.1). Moreover, every Gaussian distribution satisfies the intersection and composition properties (Studený, 2005, Proposition 2.1 and Corollary 2.4).
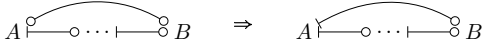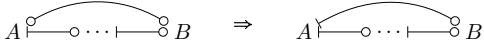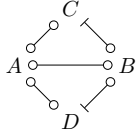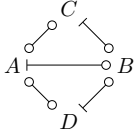
Table 1: The algorithm.

Input: A probability distribution $p$ that is faithful to an unknown CG $G$.
Output: A CG $H$ that is triplex equivalent to $G$.

1   Let $H$ denote the complete undirected graph
2   Set $l = 0$
3   Repeat while $l \leq |V| - 2$
4     For each ordered pair of nodes $A$ and $B$ in $H$ st $A \in ad_H(B)$ and $|[ad_H(A) \cup ad_H(ad_H(A))] \smallsetminus B| \geq l$
5      If there is some $S \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \smallsetminus B$ st $|S| = l$ and $A \perp_p B|S$ then
6        Set $S_{AB} = S_{BA} = S$
7        Remove the edge $A - B$ from $H$
8      Set $l = l + 1$
9   Apply the rules R1-R4 to $H$ while possible
10   Replace every edge $\vdash$ ($\mapsto$) in $H$ with $\rightarrow$ ($-$)

Table 2: The rules R1-R4.

R1:    $A \circ\!\!-\!\!-\!\!\circ B \circ\!\!-\!\!-\!\!-\!\!\circ C$    $\Rightarrow$    $A \vdash\!\!-\!\!-\!\!\circ B \circ\!\!-\!\!-\!\!\dashv C$
      $\wedge\ B \notin S_{AC}$

R2:    $A \vdash\!\!-\!\!-\!\!\circ B \circ\!\!-\!\!-\!\!-\!\!\circ C$    $\Rightarrow$    $A \vdash\!\!-\!\!-\!\!\circ B \vdash\!\!-\!\!-\!\!\circ C$
      $\wedge\ B \in S_{AC}$

R3:    $A \vdash\!\!\!\!\!\frown\!\!\cdots\!\!-\!\!\circ B$    $\Rightarrow$    $A \vdash\!\!\!\!\!\frown\!\!\cdots\!\!-\!\!\circ B$

R4:    (diagram with nodes $C$, $A$, $B$, $D$) $\Rightarrow$ (diagram with nodes $C$, $A$, $B$, $D$)
      $\wedge\ A \in S_{CD}$

rule are applied. Note that the ends of some of the edges in the rules are labeled with a circle such as in $\mapsto$ or $\circ\!\!-\!\!\circ$. The circle represents an unspecified end, i.e. a block or nothing. The modifications in the consequents of the rules consist in adding some blocks. Note that only the blocks that appear in the consequents are added, i.e. the circled ends do not get modified. The conditions in the antecedents of R1, R2 and R4 consist of an induced subgraph of $H$ and the fact that some of its nodes are or are not in some separators found in line 6. The condition in the antecedent of R3 is slightly different as it only says that there is a cycle in $H$ whose edges have certain blocks, i.e. it says nothing about the subgraph induced by the nodes in the cycle or whether these nodes belong to some separators or not. Note that, when considering the application of R3, one does not need to consider intersecting cycles, i.e. cycles containing repeated nodes other than the initial and final ones.

## 4   Correctness of the Algorithm

In this section, we prove that our algorithm is correct, i.e. it returns a CG the given probability distribution is faithful to. We start proving a result for any probability distribution that satisfies the intersection and composition properties.

Recall that any probability distribution that is faithful to a CG satisfies these properties and, thus, the following result applies to it.

**Lemma 1.** *Let $p$ denote a probability distribution that satisfies the intersection and composition properties. Then, $p$ is Markovian wrt a CG $G$ iff $p$ satisfies the following conditions:*

*C1: $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A) | pa_G(A \cup ne_G(A)) \cup ne_G(A)$ for all $A \in V$, and*

*C2: $A \perp_p V \smallsetminus A \smallsetminus de_G(A) \smallsetminus pa_G(A) | pa_G(A)$ for all $A \in V$.*

*Proof.* It follows from Andersson et al. (2001, Theorem 3) and Levitz et al. (2001, Theorem 4.1) that $p$ is Markovian wrt $G$ iff $p$ satisfies the following conditions:

L1: $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A) | [V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A))] \cup ne_G(A)$ for all $A \in V$, and

L2: $A \perp_p V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A)) \smallsetminus pa_G(A) | pa_G(A)$ for all $A \in V$.

Clearly, C2 holds iff L2 holds because $de_G(A) = [co_G(A) \cup de_G(co_G(A))] \smallsetminus A$. We prove below that if L2 holds, then C1 holds iff L1 holds. We first prove the if part.

1. $B \perp_p V \smallsetminus co_G(B) \smallsetminus de_G(co_G(B)) \smallsetminus pa_G(B)|pa_G(B)$ for all $B \in A \cup ne_G(A)$ by L2.

2. $B \perp_p V \smallsetminus co_G(B) \smallsetminus de_G(co_G(B)) \smallsetminus pa_G(A \cup ne_G(A))|pa_G(A \cup ne_G(A))$ for all $B \in A \cup ne_G(A)$ by weak union on 1.

3. $A \cup ne_G(A) \perp_p V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A)) \smallsetminus pa_G(A \cup ne_G(A))|pa_G(A \cup ne_G(A))$ by repeated application of symmetry and composition on 2.

4. $A \perp_p V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A)) \smallsetminus pa_G(A \cup ne_G(A))|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by symmetry and weak union on 3.

5. $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A)|[V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A))] \cup ne_G(A)$ by L1.

6. $A \perp_p [co_G(A) \smallsetminus A \smallsetminus ne_G(A)] \cup [V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A)) \smallsetminus pa_G(A \cup ne_G(A))]|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by contraction on 4 and 5.

7. $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A)|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by decomposition on 6.

We now prove the only if part.

8. $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A)|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by C1.

9. $A \perp_p [V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A)) \smallsetminus pa_G(A \cup ne_G(A))] \cup [co_G(A) \smallsetminus A \smallsetminus ne_G(A)]|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by composition on 4 and 8.

10. $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A)|[V \smallsetminus co_G(A) \smallsetminus de_G(co_G(A))] \cup ne_G(A)$ by weak union on 9.

$\square$

**Lemma 2.** *After line 8, $G$ and $H$ have the same adjacencies.*

*Proof.* Consider any pair of nodes $A$ and $B$ in $G$. If $A \in ad_G(B)$, then $A \not\perp_p B|S$ for all $S \subseteq V \smallsetminus [A \cup B]$ by the faithfulness assumption. Consequently, $A \in ad_H(B)$ at all times. On the other hand, if $A \notin ad_G(B)$, then consider the following cases.

**Case 1** Assume that $co_G(A) = co_G(B)$. Then, $A \perp_p co_G(A) \smallsetminus A \smallsetminus ne_G(A)|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by C1 in Lemma 1 and, thus,

$A \perp_p B|pa_G(A \cup ne_G(A)) \cup ne_G(A)$ by decomposition and $B \notin ne_G(A)$, which follows from $A \notin ad_G(B)$. Note that, as shown above, $pa_G(A \cup ne_G(A)) \cup ne_G(A) \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \smallsetminus B$ at all times.

**Case 2** Assume that $co_G(A) \neq co_G(B)$. Then, $A \notin de_G(B)$ or $B \notin de_G(A)$ because $G$ has no semidirected cycle. Assume without loss of generality that $B \notin de_G(A)$. Then, $A \perp_p V \smallsetminus A \smallsetminus de_G(A) \smallsetminus pa_G(A)|pa_G(A)$ by C2 in Lemma 1 and, thus, $A \perp_p B|pa_G(A)$ by decomposition, $B \notin de_G(A)$, and $B \notin pa_G(A)$ which follows from $A \notin ad_G(B)$. Note that, as shown above, $pa_G(A) \subseteq ad_H(A) \smallsetminus B$ at all times.

Therefore, in either case, there will exist some $S$ in line 5 such that $A \perp_p B|S$ and, thus, the edge $A - B$ will be removed from $H$ in line 7. Consequently, $A \notin ad_H(B)$ after line 8. $\square$

The next lemma proves that the rules R1-R4 are sound, i.e. if the antecedent holds in $G$, then so does the consequent.

**Lemma 3.** *The rules R1-R4 are sound.*

*Proof.* According to the antecedent of R1, $G$ has a triplex $(\{A, C\}, B)$. Then, $G$ has an induced subgraph of the form $A \to B \leftarrow C$, $A \to B - C$ or $A - B \leftarrow C$. In either case, the consequent of R1 holds.

According to the antecedent of R2, (i) $G$ does not have a triplex $(\{A, C\}, B)$, (ii) $A \to B$ or $A - B$ is in $G$, (iii) $B \in ad_G(C)$, and (iv) $A \notin ad_G(C)$. Then, $B \to C$ or $B - C$ is in $G$. In either case, the consequent of R2 holds.

According to the antecedent of R3, (i) $G$ has a descending route from $A$ to $B$, and (ii) $A \in ad_G(B)$. Then, $A \to B$ or $A - B$ is in $G$, because $G$ has no semidirected cycle. In either case, the consequent of R3 holds.

According to the antecedent of R4, neither $B \to C$ nor $B \to D$ are in $G$. Assume to the contrary that $A \leftarrow B$ is in $G$. Then, $G$ must have an induced subgraph that is consistent with

$$C$$
$$A \longleftarrow \quad \circ B$$
$$D$$
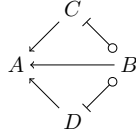
because, otherwise, it would have a semidirected cycle. However, this induced subgraph contradicts that $A \in S_{CD}$. □

**Lemma 4.** *After line 10, $G$ and $H$ have the same triplexes. Moreover, $H$ has all the immoralities in $G$.*

*Proof.* We first prove that any triplex in $H$ is in $G$. Assume to the contrary that $H$ has a triplex $(\{A,C\}, B)$ that is not in $G$. This is possible iff, when line 10 is executed, $H$ has an induced subgraph of one of the following forms:

$$A \longrightarrow B \multimap C \quad A \vdash B \multimap C \quad A \mapsto B \multimap C$$

where $B \in S_{AC}$ by Lemma 2. The first and second forms are impossible because, otherwise, $A \multimap B$ would be in $H$ by R2. The third form is impossible because, otherwise, $B \mapsto C$ would be in $H$ by R2.

We now prove that any triplex $(\{A,C\}, B)$ in $G$ is in $H$. Let the triplex be of the form $A \to B \leftarrow C$. Then, when line 10 is executed, $A \mapsto B \multimap C$ is in $H$ by R1, and neither $A \mapsto B$ nor $B \mapsto C$ is in $H$ by Lemmas 2 and 3. Then, the triplex is in $H$. Note that the triplex is an immorality in both $G$ and $H$. Likewise, let the triplex be of the form $A \to B - C$. Then, when line 10 is executed, $A \mapsto B \multimap C$ is in $H$ by R1, and $A \mapsto B$ is not in $H$ by Lemmas 2 and 3. Then, the triplex is in $H$. Note that the triplex is a flag in $G$ but it may be an immorality in $H$. □

**Lemma 5.** *After line 9, $H$ does not have any induced subgraph of the form* $A \mapsto B \multimap C$.

*Proof.* Assume to the contrary that the lemma does not hold. Consider the following cases.

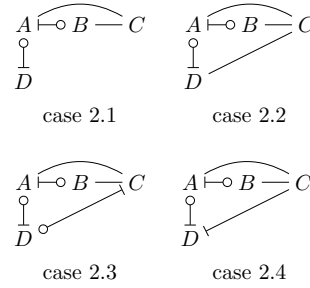**Case 1** Assume that $A \mapsto B$ is in $H$ due to R1. Then, when R1 was applied, $H$ had an induced subgraph of one of the following forms:

$$A \mapsto B \multimap C \qquad A \mapsto B \multimap C$$
$$\quad | \qquad\qquad\qquad |$$
$$D \qquad\qquad\qquad D$$
case 1.1 · case 1.2
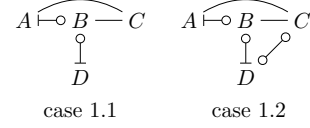
**Case 1.1** If $B \notin S_{CD}$ then $B \multimap C$ is in $H$ by R1, else $B \vdash C$ is in $H$ by R2. Either case is a contradiction.

**Case 1.2** If $C \notin S_{AD}$ then $A \vdash C$ is in $H$ by R1, else $B \multimap C$ is in $H$ by R4. Either case is a contradiction.

**Case 2** Assume that $A \mapsto B$ is in $H$ due to R2. Then, when R2 was applied, $H$ had an induced subgraph of one of the following forms:

case 2.1 · case 2.2 · case 2.3 · case 2.4

**Case 2.1** If $A \notin S_{CD}$ then $A \multimap C$ is in $H$ by R1, else $A \vdash C$ is in $H$ by R2. Either case is a contradiction.

**Case 2.2** Restart the proof with $D$ instead of $A$ and $A$ instead of $B$.

**Case 2.3** Then, $A \multimap C$ is in $H$ by R3, which is a contradiction.

**Case 2.4** If $C \notin S_{BD}$ then $B \vdash C$ is in $H$ by R1, else $B \multimap C$ is in $H$ by R2. Either case is a contradiction.

**Case 3** Assume that $A \mapsto B$ is in $H$ due to R3. Then, when R3 was applied, $H$ had an induced subgraph of one of the following forms:

case 3.1 · case 3.2 · case 3.3 · case 3.4
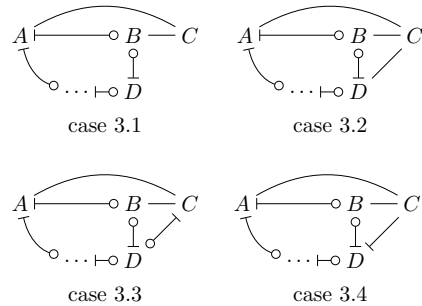
Note that $C$ cannot belong to the route $A \mapsto \ldots \mapsto D$ because, otherwise, $A \mapsto C$ would be in $H$ by R3.
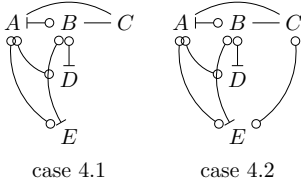
**Case 3.1** If $B \notin S_{CD}$ then $B \dashv C$ is in $H$ by R1, else $B \vdash C$ is in $H$ by R2. Either case is a contradiction.

**Case 3.2** Restart the proof with $D$ instead of $A$.

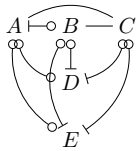**Case 3.3** Then, $B \dashv C$ is in $H$ by R3, which is a contradiction.

**Case 3.4** Then, $A \vdash C$ is in $H$ by R3, which is a contradiction.

**Case 4** Assume that $A \mapsto B$ is in $H$ due to R4. Then, when R4 was applied, $H$ had an induced subgraph of one of the following forms:



case 4.1          case 4.2



case 4.3          case 4.4
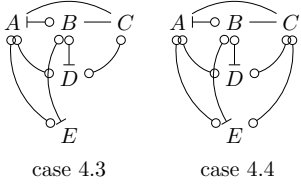
**Cases 4.1-4.3** If $B \notin S_{CD}$ or $B \notin S_{CE}$ then $B \dashv C$ is in $H$ by R1, else $B \vdash C$ is in $H$ by R2. Either case is a contradiction.

**Case 4.4** Assume that $C \in S_{DE}$. Then, $B \dashv C$ is in $H$ by R4, which is a contradiction. On the other hand, assume that $C \notin S_{DE}$. Then, it follows from applying R1 that $H$ has an induced subgraph of the form



Note that $A \in S_{DE}$ because, otherwise, R4 would not have been applied.

Then, $A \vdash C$ is in $H$ by R4, which is a contradiction.

$\square$

**Lemma 6.** *After line 9, every cycle in $H$ that has an edge $\vdash$ also has an edge $\dashv$.*

*Proof.* Assume to the contrary that $H$ has a cycle $\rho : V_1, \ldots, V_n = V_1$ that has an edge $\vdash$ but no edge $\dashv$. Note that every edge in $\rho$ cannot be $\mapsto$ because, otherwise, every edge in $\rho$ would be $\mapsto$ by repeated application of R3, which contradicts the assumption that $\rho$ has an edge $\vdash$. Therefore, $\rho$ has an edge $-$ or $\dashv$. Since the latter contradicts the assumption that the lemma does not hold, $\rho$ has an edge $-$. Assume that $\rho$ is of length three. Then, $\rho$ is of one of the following forms:



The first form is impossible by Lemma 5. The second form is impossible because, otherwise, $V_2 \dashv V_3$ would be in $H$ by R3. The third form is impossible because, otherwise, $V_1 \vdash V_3$ would be in $H$ by R3. Thus, the lemma holds for cycles of length three.

Assume that $\rho$ is of length greater than three. Recall from above that $\rho$ has an edge $-$ and no edge $\dashv$. Let $V_{i+1} - V_{i+2}$ be the first edge $-$ in $\rho$. Assume without loss of generality that $i > 0$. Then, $\rho$ has a subpath of the form $V_i \mapsto V_{i+1} - V_{i+2}$. Note that $V_i \in ad_H(V_{i+2})$ because, otherwise, if $V_{i+1} \notin S_{V_i V_{i+2}}$ then $V_{i+1} \dashv V_{i+2}$ would be in $H$ by R1, else $V_{i+1} \vdash V_{i+2}$ would be in $H$ by R2. Thus, $H$ has an induced subgraph of one of the following forms:



The first form is impossible by Lemma 5. The second form is impossible because, otherwise, $V_{i+1} \dashv V_{i+2}$ would be in $H$ by R3. Thus, the third form is the only possible. Note that this implies that $\varrho : V_1, \ldots, V_i, V_{i+2}, \ldots, V_n = V_1$ is a cycle in $H$ that has an edge $\vdash$ and no edge $\dashv$.

By repeatedly applying the reasoning above, one can see that $H$ has a cycle of length three

that has an edge ⊢ and no edge ⊣. As shown above, this is impossible. Thus, the lemma holds for cycles of length greater than three too. □

**Theorem 1.** *After line 10, $H$ is triplex equivalent to $G$ and it has no semidirected cycle.*

*Proof.* Lemma 2 implies that $G$ and $H$ have the same adjacencies. Lemma 4 implies that $G$ and $H$ have the same triplexes. Lemma 6 implies that $H$ has no semidirected cycle. □

## 5 Discussion

In this paper, we have presented an algorithm for learning an AMP CG a given probability distribution $p$ is faithful to. In practice, of course, we do not usually have access to $p$ but to a finite sample from it. Our algorithm can easily be modified to deal with this situation: Replace $A \perp_p B | S$ in line 5 with a hypothesis test, preferably with one that is consistent so that the resulting algorithm is asymptotically correct.

It is worth mentioning that, whereas R1, R2 and R4 only involve three or four nodes, R3 may involve many more. Hence, it would be desirable to replace R3 with a simpler rule such as

$$A \overset{\frown}{\vdash} \circ\, B \vdash\!\circ\, C \quad \Rightarrow \quad A \overset{\frown}{\vdash} \circ\, B \vdash\!\circ\, C \,.$$

Unfortunately, we have not succeeded so far in proving the correctness of our algorithm with such a simpler rule. Note that the output of our algorithm will be the same whether we keep R3 or we replace it with a simpler sound rule. The only benefit of the simpler rule may be a decrease in running time.
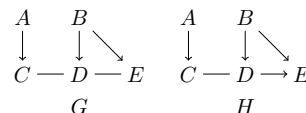
We have shown in Lemma 4 that, after line 10, $H$ has all the immoralities in $G$ or, in other words, every flag in $H$ is in $G$. The following lemma strengthens this fact.

**Lemma 7.** *After line 10, every flag in $H$ is in every CG $F$ that is triplex equivalent to $G$.*

*Proof.* Note that every flag in $H$ is due to an induced subgraph of the form $A \vdash B \dashv C$. Note also that all the blocks in $H$ follow from the

adjacencies and triplexes in $G$ by repeated application of R1-R4. Since $G$ and $F$ have the same adjacencies and triplexes, all the blocks in $H$ hold in both $G$ and $F$ by Lemma 3. □
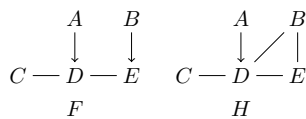
The lemma above implies that, in terms of Roverato and Studený (2006), our algorithm outputs a deflagged graph. Roverato and Studený (2006) also introduce the concept of strongly equivalent CGs: Two CGs are strongly equivalent iff they have the same adjacencies, immoralities and flags. Unfortunately, not every edge → in $H$ after line 10 is in every deflagged graph that is triplex equivalent to $G$, as the following example illustrates, where both $G$ and $H$ are deflagged graphs.

$$
\begin{array}{ccc}
A & B & \\
\downarrow & \downarrow & \searrow \\
C \!-\! D & \!-\! E &
\end{array}
\qquad
\begin{array}{ccc}
A & B & \\
\downarrow & \downarrow & \searrow \\
C \!-\! D & \!\rightarrow\! E &
\end{array}
$$
$$\phantom{xxx}G\phantom{xxxxxxxxx}H$$

Therefore, in terms of Roverato and Studený (2006), our algorithm outputs a deflagged graph but not the largest deflagged graph. The latter is a distinguished member of a class of triplex equivalent CGs. Fortunately, the largest deflagged graph can easily be obtained from any deflagged graph in the class (Roverato and Studený, 2006, Corollary 17).

The correctness of our algorithm lies upon the assumption that $p$ is faithful to some CG. This is a strong requirement that we would like to weaken, e.g. by replacing it with the milder assumption that $p$ satisfies the composition property. Correct algorithms for learning directed and acyclic graphs (a.k.a Bayesian networks) under the composition property assumption exist (Chickering and Meek, 2002; Nielsen et al., 2003). We have recently developed a correct algorithm for learning LWF CGs under the composition property (Peña et al., 2012). The way in which these algorithms proceed (a.k.a. score+search based approach) is rather different from that of the algorithm presented in this paper (a.k.a. constraint based approach). In a nutshell, they can be seen as consisting of two phases: A first phase that starts from the empty graph $H$ and adds single edges to

it until $p$ is Markovian wrt $H$, and a second phase that removes single edges from $H$ until $p$ is Markovian wrt $H$ and $p$ is not Markovian wrt any CG $F$ st $I(H) \subseteq I(F)$. The success of the first phase is guaranteed by the composition property assumption, whereas the success of the second phase is guaranteed by the so-called Meek's conjecture (Meek, 1997). Specifically, given two directed and acyclic graphs $F$ and $H$ st $I(H) \subseteq I(F)$, Meek's conjecture states that we can transform $F$ into $H$ by a sequence of operations st, after each operation, $F$ is a directed and acyclic graph and $I(H) \subseteq I(F)$. The operations consist in adding a single edge to $F$, or replacing $F$ with a triplex equivalent directed and acyclic graph. Meek's conjecture was proven to be true in (Chickering, 2002, Theorem 4). The extension of Meek's conjecture to LWF CGs was proven to be true in (Peña, 2011, Theorem 1). Unfortunately, the extension of Meek's conjecture to AMP CGs does not hold, as the following example illustrates.

$$\begin{array}{cc} \begin{array}{ccc} A & B & \\ \downarrow & \downarrow & \\ C\!-\!\!D\!-\!\!E \\ F \end{array} & \begin{array}{ccc} A & B & \\ \downarrow\!\nearrow & \downarrow & \\ C\!-\!\!D\!-\!\!E \\ H \end{array} \end{array}$$

Then, $I(H) = \{X \perp_H Y|Z : X \perp_H Y|Z \in I_1(H) \cup I_2(H) \vee Y \perp_H X|Z \in I_1(H) \cup I_2(H)\}$ where $I_1(H) = \{A \perp_H Y|Z : Y, Z \subseteq B \cup C \cup E \wedge D \notin Z\}$ and $I_2(H) = \{C \perp_H Y|Z : Y \subseteq B \cup E \wedge A \cup D \subseteq Z\}$. One can easily confirm that $I(H) \subseteq I(F)$ by using the definition of separation. However, there is no CG that is triplex equivalent to $F$ or $H$ and, obviously, one cannot transform $F$ into $H$ by adding a single edge.

While the example above compromises the development of score+search learning algorithms that are correct and efficient under the composition property assumption, it is not clear to us whether it also does it for constraint based algorithms. This is something we plan to study.

## Acknowledgments

## References

Andersson, S. A., Madigan, D. and Perlman, M. D. Alternative Markov Properties for Chain Graphs. *Scandinavian Journal of Statistics*, 28:33-85, 2001.

Chickering, D. M. Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research*, 3:507-554, 2002.

Chickering, D. M. and Meek, C. Finding Optimal Bayesian Networks. In *Proceedings of 18th Conference on Uncertainty in Artificial Intelligence*, 94-102, 2002.

Drton, M. and Eichler, M. Maximum Likelihood Estimation in Gaussian Chain Graph Models under the Alternative Markov Property. *Scandinavian Journal of Statistics*, 33:247-257, 2006.

Lauritzen, S. L. *Graphical Models*. Oxford University Press, 1996.

Levitz, M., Perlman M. D. and Madigan, D. Separation and Completeness Properties for AMP Chain Graph Markov Models. *The Annals of Statistics*, 29:1751-1784, 2001.

Ma, Z., Xie, X. and Geng, Z. Structural Learning of Chain Graphs via Decomposition. *Journal of Machine Learning Research*, 9:2847-2880, 2008.

Meek, C. Causal Inference and Causal Explanation with Background Knowledge. *Proceedings of 11th Conference on Uncertainty in Artificial Intelligence*, 403-418, 1995.

Meek, C. *Graphical Models: Selecting Causal and Statistical Models*. PhD thesis, Carnegie Mellon University, 1997.

Nielsen, J. D., Kočka, T. and Peña, J. M. On Local Optima in Learning Bayesian Networks. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, 435-442, 2003.

Peña, J. M. Towards Optimal Learning of Chain Graphs. arXiv:1109.5404v1 [stat.ML], 2011.

Peña, J. M., Sonntag, D. and Nielsen, J. D. An Inclusion Optimal Algorithm for Chain Graph Structure Learning. Submitted, 2012.

Roverato, A. and Studený, M. A Graphical Representation of Equivalence Classes of AMP Chain Graphs. *Journal of Machine Learning Research*, 7:1045-1078, 2006.

Spirtes, P., Glymour, C. and Scheines, R. *Causation, Prediction, and Search*. Springer-Verlag, 1993.

Studený, M. A Recovery Algorithm for Chain Graphs. *International Journal of Approximate Reasoning*, 17:265-293, 1997.

Studený, M. *Probabilistic Conditional Independence Structures*. Springer, 2005.