

# Learning Marginal AMP Chain Graphs under Faithfulness

Jose M. Peña

ADIT, IDA, Linköping University, SE-58183 Linköping, Sweden  
jose.m.pena@liu.se

**Abstract.** Marginal AMP chain graphs are a recently introduced family of models that is based on graphs that may have undirected, directed and bidirected edges. They unify and generalize the AMP and the multivariate regression interpretations of chain graphs. In this paper, we present a constraint based algorithm for learning a marginal AMP chain graph from a probability distribution which is faithful to it. We also show that the extension of Meek’s conjecture to marginal AMP chain graphs does not hold, which compromises the development of efficient and correct score+search learning algorithms under assumptions weaker than faithfulness.

**Keywords:** Chain graphs, marginal AMP chain graphs, learning graphical models

## 1 Introduction

Chain graphs (CGs) are graphs with possibly directed and undirected edges, and no semidirected cycle. They have been extensively studied as a formalism to represent independence models, because they can model symmetric and asymmetric relationships between the random variables of interest. However, there are three different interpretations of CGs as independence models: The Lauritzen-Wermuth-Frydenberg (LWF) interpretation [6], the multivariate regression (MVR) interpretation [4], and the Andersson-Madigan-Perlman (AMP) interpretation [1]. It is worth mentioning that no interpretation subsumes another: There are many independence models that can be represented by a CG under one interpretation but that cannot be represented by any CG under the other interpretations [1, 17]. Moreover, although MVR CGs were originally represented using dashed directed and undirected edges, we like other authors prefer to represent them using solid directed and bidirected edges.

Recently, a new family of models has been proposed to unify and generalize the AMP and MVR interpretations of CGs [11]. This new family, named marginal AMP (MAMP) CGs, is based on graphs that may have undirected, directed and bidirected edges. In this paper, we extend [11] by presenting an algorithm for learning an MAMP CG from a probability distribution which is faithful to it. Our algorithm is constraint based and builds upon those developed in [16] and [10] for learning, respectively, MVR and AMP CGs under the

faithfulness assumption. Finally, note that there also exist algorithms for learning LWF CGs under the faithfulness assumption [7, 19] and under the milder composition property assumption [13]. In this paper, we also show that the extension of Meek’s conjecture to MAMP CGs does not hold, which compromises the development of efficient and correct score+search learning algorithms under assumptions weaker than faithfulness.

The rest of this paper is organized as follows. We start with some preliminaries in Section 2. Then, we introduce MAMP CGs in Section 3, followed by the algorithm for learning them in Section 4. We close the paper with some discussion in Section 5.

## 2 Preliminaries

In this section, we introduce some concepts of models based on graphs, i.e. graphical models. Most of these concepts have a unique definition in the literature. However, a few concepts have more than one and we opt for the most suitable in this work. All the graphs and probability distributions in this paper are defined over a finite set  $V$ . All the graphs in this paper are simple, i.e. they contain at most one edge between any pair of nodes. The elements of  $V$  are not distinguished from singletons.

If a graph  $G$  contains an undirected, directed or bidirected edge between two nodes  $V_1$  and  $V_2$ , then we write that  $V_1 - V_2$ ,  $V_1 \rightarrow V_2$  or  $V_1 \leftrightarrow V_2$  is in  $G$ . We represent with a circle, such as in  $\circ\rightarrow$  or  $\circ\leftrightarrow$ , that the end of an edge is unspecified, i.e. it may be an arrow tip or nothing. 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 children of  $X$  is the set  $ch_G(X) = \{V_1 | V_1 \leftarrow V_2 \text{ is in } G, V_1 \notin X \text{ and } V_2 \in X\}$ . The neighbors of  $X$  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 spouses of  $X$  is the set  $sp_G(X) = \{V_1 | V_1 \leftrightarrow V_2 \text{ is in } G, V_1 \notin X \text{ and } V_2 \in X\}$ . The adjacents of  $X$  is the set  $ad_G(X) = ne_G(X) \cup pa_G(X) \cup ch_G(X) \cup sp_G(X)$ . A route between a node  $V_1$  and a node  $V_n$  in  $G$  is a sequence of (not necessarily distinct) nodes  $V_1, \dots, V_n$  such that  $V_i \in ad_G(V_{i+1})$  for all  $1 \leq i < n$ . If the nodes in the route are all distinct, then the route is called a path. The length of a route is the number of (not necessarily distinct) edges in the route, e.g. the length of the route  $V_1, \dots, V_n$  is  $n - 1$ . A route is called descending if  $V_i \rightarrow V_{i+1}$  or  $V_i - V_{i+1}$  is in  $G$  for all  $1 \leq i < n$ . A route is called strictly descending if  $V_i \rightarrow V_{i+1}$  is in  $G$  for all  $1 \leq i < n$ . The descendants of a set of nodes  $X$  of  $G$  is the set  $de_G(X) = \{V_n | \text{there is a descending route from } V_1 \text{ to } V_n \text{ in } G, V_1 \in X \text{ and } V_n \notin X\}$ . The strict ascendants of  $X$  is the set  $san_G(X) = \{V_1 | \text{there is a strictly descending route from } V_1 \text{ to } V_n \text{ in } G, V_1 \notin X \text{ and } V_n \in X\}$ . A route  $V_1, \dots, V_n$  in  $G$  is called a cycle if  $V_n = V_1$ . Moreover, it is called a semidirected cycle if  $V_n = V_1$ ,  $V_1 \rightarrow V_2$  is in  $G$  and  $V_i \rightarrow V_{i+1}$ ,  $V_i \leftrightarrow V_{i+1}$  or  $V_i - V_{i+1}$  is in  $G$  for all  $1 < i < n$ . A cycle has a chord if two non-consecutive nodes of the cycle are adjacent in  $G$ . An AMP chain graph (AMP CG) is a graph whose every edge is directed or undirected such that it has no semidirected cycles. A MVR chain graph (MVR CG) is a graph whose every edge is directed or bidirected such that it has no semidirected

cycles. 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  $X$ .

We now recall the semantics of AMP and MVR CGs. A node  $B$  in a path  $\rho$  in an AMP CG  $G$  is called a triplex node in  $\rho$  if  $A \rightarrow B \leftarrow C$ ,  $A \rightarrow B - C$ , or  $A - B \leftarrow C$  is a subpath of  $\rho$ . Moreover,  $\rho$  is said to be  $Z$ -open with  $Z \subseteq V$  when

- every triplex node in  $\rho$  is in  $Z \cup \text{san}_G(Z)$ , and
- every non-triplex node  $B$  in  $\rho$  is outside  $Z$ , unless  $A - B - C$  is a subpath of  $\rho$  and  $pa_G(B) \setminus Z \neq \emptyset$ .

A node  $B$  in a path  $\rho$  in an MVR CG  $G$  is called a triplex node in  $\rho$  if  $A \leftrightarrow B \leftarrow C$  is a subpath of  $\rho$ . Moreover,  $\rho$  is said to be  $Z$ -open with  $Z \subseteq V$  when

- every triplex node in  $\rho$  is in  $Z \cup \text{san}_G(Z)$ , and
- every non-triplex node  $B$  in  $\rho$  is outside  $Z$ .

Let  $X, Y$  and  $Z$  denote three disjoint subsets of  $V$ . When there is no  $Z$ -open path in an AMP or MVR CG  $G$  between a node in  $X$  and a node in  $Y$ , we say that  $X$  is separated from  $Y$  given  $Z$  in  $G$  and denote it as  $X \perp_G Y | Z$ . The independence model represented by  $G$ , denoted as  $I(G)$ , is the set of separations  $X \perp_G Y | Z$ . In general,  $I(G)$  is different whether  $G$  is an AMP or MVR CG.

### 3 MAMP CGs

In this section, we review marginal AMP (MAMP) CGs. We refer the reader to [11] for more details. Specifically, a graph  $G$  containing possibly directed, bidirected and undirected edges is an MAMP CG if

- C1.**  $G$  has no semidirected cycle,
- C2.**  $G$  has no cycle  $V_1, \dots, V_n = V_1$  such that  $V_1 \leftrightarrow V_2$  is in  $G$  and  $V_i - V_{i+1}$  is in  $G$  for all  $1 < i < n$ , and
- C3.** if  $V_1 - V_2 - V_3$  is in  $G$  and  $sp_G(V_2) \neq \emptyset$ , then  $V_1 - V_3$  is in  $G$  too.

The semantics of MAMP CGs is as follows. A node  $B$  in a path  $\rho$  in an MAMP CG  $G$  is called a triplex node in  $\rho$  if  $A \leftrightarrow B \leftarrow C$ ,  $A \leftrightarrow B - C$ , or  $A - B \leftarrow C$  is a subpath of  $\rho$ . Moreover,  $\rho$  is said to be  $Z$ -open with  $Z \subseteq V$  when

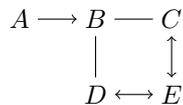
- every triplex node in  $\rho$  is in  $Z \cup \text{san}_G(Z)$ , and
- every non-triplex node  $B$  in  $\rho$  is outside  $Z$ , unless  $A - B - C$  is a subpath of  $\rho$  and  $sp_G(B) \neq \emptyset$  or  $pa_G(B) \setminus Z \neq \emptyset$ .

Let  $X, Y$  and  $Z$  denote three disjoint subsets of  $V$ . When there is no  $Z$ -open path in  $G$  between a node in  $X$  and a node in  $Y$ , we say that  $X$  is separated from  $Y$  given  $Z$  in  $G$  and denote it as  $X \perp_G Y | Z$ . The independence model represented by  $G$ , denoted as  $I(G)$ , is the set of separations  $X \perp_G Y | Z$ . We denote by  $X \perp_p Y | Z$  (respectively  $X \not\perp_p Y | Z$ ) that  $X$  is independent (respectively

dependent) of  $Y$  given  $Z$  in a probability distribution  $p$ . 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  $V$ . We say that two MAMP CGs are Markov equivalent if they represent the same independence model. In an MAMP CG, a triplex  $(\{A, C\}, B)$  is an induced subgraph of the form  $A \leftrightarrow B \leftrightarrow C$ ,  $A \leftrightarrow B - C$ , or  $A - B \leftrightarrow C$ . We say that two MAMP CGs are triplex equivalent if they have the same adjacencies and the same triplexes. Two MAMP CGs are Markov equivalent iff they are triplex equivalent [11, Theorem 7].

Clearly, the union of AMP and MVR CGs is a subfamily of MAMP CGs. The following example shows that it is a proper subfamily.

*Example 1.* The independence model represented by the MAMP CG  $G$  below cannot be represented by any AMP or MVR CG.



To see it, assume to the contrary that it can be represented by an AMP CG  $H$ . Note that  $H$  is an MAMP CG too. Then,  $G$  and  $H$  must have the same triplexes. Then,  $H$  must have triplexes  $(\{A, D\}, B)$  and  $(\{A, C\}, B)$  but no triplex  $(\{C, D\}, B)$ . So,  $C - B - D$  must be in  $H$ . Moreover,  $H$  must have a triplex  $(\{B, E\}, C)$ . So,  $C \leftarrow E$  must be in  $H$ . However, this implies that  $H$  does not have a triplex  $(\{C, D\}, E)$ , which is a contradiction because  $G$  has such a triplex. To see that no MVR CG can represent the independence model represented by  $G$ , simply note that no MVR CG can have triplexes  $(\{A, D\}, B)$  and  $(\{A, C\}, B)$  but no triplex  $(\{C, D\}, B)$ .

Finally, other families of models that are based on graphs that may contain undirected, directed and bidirected edges are summary graphs after replacing the dashed undirected edges with bidirected edges [4], MC graphs [5], maximal ancestral graphs [14], and loopless mixed graphs [15]. However, the separation criteria for these families are identical to that of MVR CGs. Then, MVR CGs are a subfamily of these families but AMP CGs are not. See also [14, p. 1025] and [15, Sections 4.1-4.3]. Therefore, MAMP CGs are the only graphical models in the literature that generalize both AMP and MVR CGs.

## 4 Algorithm for Learning MAMP CGs

In this section, we present our algorithm for learning an MAMP CG from a probability distribution which is faithful to it. The algorithm builds upon those developed in [16] and [10] for learning, respectively, MVR and AMP CGs under the faithfulness assumption. The algorithm, which can be seen in Table 1, resembles the well-known PC algorithm developed in [18] for learning Bayesian networks under the faithfulness assumption, in the sense that it consists of two

**Table 1.** Algorithm for learning MAMP CGs.

---

Input: A probability distribution  $p$  that is faithful to an unknown MAMP CG  $G$ .  
Output: An MAMP 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  $[[ad_H(A) \cup ad_H(ad_H(A))] \setminus \{A, B\}] \geq l$
- 5 If there is some  $S \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \setminus \{A, B\}$  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  $A - B$  in  $H$  with  $A \rightarrow B$
- 11 Replace every edge  $A - B$  or  $A \rightarrow B$  in  $H$  with  $A \leftrightarrow B$
- 12 Replace every induced subgraph  $A \leftrightarrow B \leftrightarrow C$  in  $H$  st  $B \in S_{AC}$  with  $A - B - C$
- 13 If  $H$  has an induced subgraph  $A \overset{\curvearrowright}{\leftrightarrow} B \overset{\curvearrowleft}{\leftrightarrow} C$  then
- 14 Replace the edge  $A \leftrightarrow B$  in  $H$  with  $A - B$
- 15 Go to line 13
- 16 Return  $H$

---

**Table 2.** Rules R1-R4 in the algorithm for learning MAMP CGs.

---

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

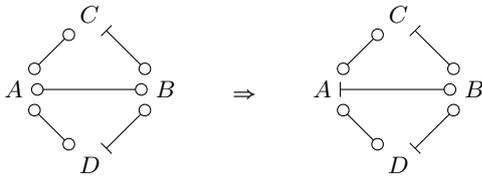
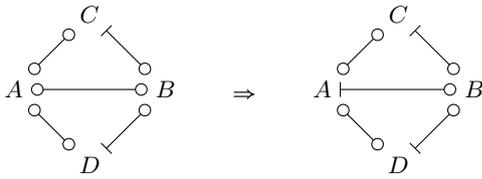
---

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

---

R3:  $A \overset{\curvearrowright}{\leftrightarrow} \dots \vdash \overset{\curvearrowleft}{\leftrightarrow} B \Rightarrow A \overset{\curvearrowright}{\vdash} \dots \vdash \overset{\curvearrowleft}{\vdash} B$

---

R4:   $\Rightarrow$    
 $\wedge A \in S_{CD}$

---

phases: The first phase (lines 1-8) aims at learning adjacencies, whereas the second phase (lines 9-15) aims at directing some of the adjacencies learnt. Specifically, the first phase declares that two nodes are adjacent if and only if 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 (line 9), 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 at the edge end such as in  $\perp$  or  $\dashv$ , and it means that the edge cannot be a directed edge pointing in the direction of the block. Note that  $\dashv$  does not mean that the edge must be undirected: It means that the edge cannot be a directed edge in either direction and, thus, it must be a bidirected or undirected edge. In the second step (lines 10-15), some edges get directed. Specifically, the edges with exactly one unblocked end get directed in the direction of the unblocked end. The rest of the edges get bidirected (see line 11), unless this produces a false triplex (see line 12) or violates the constraint C2 (see lines 13-15). Note that only cycles of length three are checked for the violation of the constraint C2.

The rules R1-R4 in Table 2 work as follows: If the conditions in the antecedent of a rule are satisfied, then the modifications in the consequent of the rule are applied. Note that the ends of some of the edges in the rules are labeled with a circle such as in  $\circ$  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 consists of just an induced subgraph of  $H$ . Specifically, the antecedent says that there is a cycle in  $H$  whose edges have certain blocks. Note that the cycle must be chordless.

The rest of this section is devoted to prove that our algorithm is correct, i.e. it returns an MAMP CG the given probability distribution is faithful to. We start by proving some auxiliary results.

**Lemma 1.** *After having executed 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  $S \subseteq V \setminus \{A, B\}$  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  $A \notin de_G(B)$  or  $B \notin de_G(A)$ . Assume without loss of generality that  $B \notin de_G(A)$ . Then,  $A \perp_p B | pa_G(A)$  [11, Theorem 5]. Note that, as shown above,  $pa_G(A) \subseteq ad_H(A) \setminus B$  at all times.
- Case 2** Assume that  $A \in de_G(B)$  and  $B \in de_G(A)$ . Then,  $A \perp_p B | ne_G(A) \cup pa_G(A \cup ne_G(A))$  [11, Theorem 5]. Note that, as shown above,  $ne_G(A) \cup pa_G(A \cup ne_G(A)) \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \setminus \{A, B\}$  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.

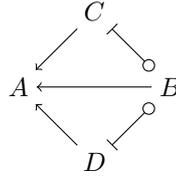
**Lemma 2.** *The rules R1-R4 block the end of an edge only if the edge is not a directed edge in  $G$  pointing in the direction of the block.*

*Proof.* According to the antecedent of R1,  $G$  has a triplex  $(\{A, C\}, B)$ . Then,  $G$  has an induced subgraph of the form  $A \leftrightarrow B \leftrightarrow C$ ,  $A \leftrightarrow B - C$  or  $A - B \leftrightarrow C$ . 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)  $A \leftrightarrow B$  or  $A - B$  is in  $G$ , (iii)  $B \in ad_G(C)$ , and (iv)  $A \notin ad_G(C)$ . Then,  $B \rightarrow C$  or  $B - C$  is in  $G$ . In either case, the consequent of R2 holds.

According to the antecedent of R3, (i)  $G$  has a path from  $A$  to  $B$  with no directed edge pointing in the direction of  $A$ , and (ii)  $A \in ad_G(B)$ . Then,  $A \leftarrow B$  cannot be in  $G$  because  $G$  has no semidirected cycle. Then, the consequent of R3 holds.

According to the antecedent of R4, neither  $B \rightarrow C$  nor  $B \rightarrow D$  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



because, otherwise,  $G$  has a semidirected cycle. However, this contradicts that  $A \in S_{CD}$ .

**Lemma 3.** *At line 16, all the undirected edges in  $H$  are in  $G$ .*

*Proof.* Note that lines 10-11 imply that  $H$  has no undirected edge when line 12 is to be executed. Note also that any undirected edges  $A - B$  and  $B - C$  added to  $H$  in line 12 must exist in  $G$ , because this implies that  $H$  has an induced subgraph  $A \dashv B \dashv C$  with  $B \in S_{AC}$  when line 11 is to be executed, which implies that (i)  $A$  and  $B$  as well as  $B$  and  $C$  are adjacent in  $G$  whereas  $A$  and  $C$  are not adjacent in  $G$  by Lemma 1, and (ii)  $G$  has no directed edge between  $A$  and  $B$  or  $B$  and  $C$ . Then,  $A - B - C$  must be in  $G$  by Lemma 2 and the fact that  $B \in S_{AC}$ .

The paragraph above implies that all the undirected edges in  $H$  are in  $G$  when lines 13-15 are to be executed for the first time, which implies that the undirected edge added to  $H$  in the first execution of lines 13-15 must also be in  $G$  due to the constraints C1 and C2. By repeatedly applying this argument, all the undirected edges in  $H$  at line 16 must be in  $G$ .

**Lemma 4.** *At line 16,  $G$  and  $H$  have the same triplexes.*

*Proof.* We first prove that any triplex in  $H$  at line 16 is in  $G$ . Assume to the contrary that  $H$  at line 16 has a triplex  $(\{A, C\}, B)$  that is not in  $G$ . This is possible if and only if  $H$  has an induced subgraph of one of the following forms when lines 10-11 are to be executed:

$$\begin{array}{l} A \text{ --- } B \text{ --- } C \quad A \text{ --- } B \text{ } \circ \text{---} C \quad A \text{ ---} B \text{ ---} C \\ A \text{ ---} B \text{ } \circ \text{---} C \quad A \text{ ---} B \text{ ---} \circ C \quad A \text{ ---} B \text{ ---} C \end{array}$$

Note that the induced subgraphs above together with Lemma 1 imply that  $A$  is adjacent to  $B$  in  $G$ ,  $B$  is adjacent to  $C$  in  $G$ , and  $A$  is not adjacent to  $C$  in  $G$ . This together with the assumption made above that  $G$  has no triplex  $(\{A, C\}, B)$  implies that  $B \in S_{AC}$ . Now, note that the second and fourth induced subgraphs above are impossible because, otherwise,  $A \text{ } \circ \text{---} B$  would be in  $H$  by R2. Likewise, the third and fifth induced subgraphs above are impossible because, otherwise,  $B \text{ ---} C$  would be in  $H$  by R2. Now, note that any triplex that is added to  $H$  in line 11 due to the first and sixth induced subgraphs above is removed from  $H$  in line 12 because, as shown above,  $B \in S_{AC}$ . Finally, note that no triplex is added to  $H$  in lines 13-15.

We now prove that any triplex  $(\{A, C\}, B)$  in  $G$  is in  $H$  at line 16. Note that  $B \notin S_{AC}$ . Consider the following cases.

**Case 1** Assume that the triplex in  $G$  is of the form  $A \rightarrow B \circ \rightarrow C$  (respectively  $A \circ \rightarrow B \leftarrow C$ ). Then, when lines 10-11 are to be executed,  $A \text{ ---} B \text{ } \circ \text{---} C$  (respectively  $A \text{ ---} B \text{ ---} C$ ) is in  $H$  by R1 and Lemmas 1 and 2. Then, the triplex is added to  $H$  in lines 10-11. Moreover, the triplex added is of the form  $A \rightarrow B \circ \rightarrow C$  (respectively  $A \circ \rightarrow B \leftarrow C$ ) and, thus, it does not get removed from  $H$  in lines 12-15.

**Case 2** Assume that the triplex in  $G$  is of the form  $A \leftrightarrow B \circ \leftrightarrow C$  or  $A \circ \leftrightarrow B \leftrightarrow C$ . Then, when lines 10-11 are to be executed,  $A \text{ ---} B \text{ } \circ \text{---} C$  is in  $H$  by R1 and Lemmas 1 and 2. Then, the triplex is added to  $H$  in lines 10-11. Moreover, the triplex cannot get removed from  $H$  in lines 12-15. To see it, assume the contrary. Note that all lines 12-15 do is replacing bidirected edges in  $H$  with undirected edges. Thus, the triplex cannot get removed from  $H$  unless it is of the form  $A \leftrightarrow B \leftrightarrow C$ ,  $A \leftrightarrow B - C$ , or  $A - B \leftrightarrow C$ . Consider the following cases.

**Case 2.1** Assume that the triplex gets removed from  $H$  in line 12. Assume that the triplex is of the form  $A \leftrightarrow B \leftrightarrow C$ . The proofs for the forms  $A \leftrightarrow B - C$  and  $A - B \leftrightarrow C$  are similar. Note that the triplex cannot get removed from  $H$  by applying line 12 to  $A \leftrightarrow B \leftrightarrow C$  because, as shown above,  $B \notin S_{AC}$ . Then, for the triplex to get removed from  $H$  in line 12,  $H$  must have two induced subgraphs  $A' \leftrightarrow A \leftrightarrow B$  and  $B \leftrightarrow C \leftrightarrow C'$  with  $A \in S_{A'B}$  and  $C \in S_{BC'}$  when line 12 is to be executed. This implies that  $A \text{ ---} B \text{ ---} C$  is in  $H$  when lines 10-11 are to be executed because, as shown above,  $A \text{ ---} B \text{ } \circ \text{---} C$  is in  $H$  when lines 10-11 are to be executed. Therefore,  $H$  has two induced subgraphs  $A' \text{ ---} A \text{ ---} B$  and  $B \text{ ---} C \text{ ---} C'$  by R2 when lines 10-11 are to be executed. Then,  $A' - A$  or  $A' \leftrightarrow A$  must

be in  $G$  by Lemmas 1 and 2. Then,  $A - B$  or  $A \rightarrow B$  must be in  $G$  by Lemmas 1 and 2 and the fact that  $A \in S_{A'B}$ . However,  $A \rightarrow B$  cannot be in  $G$  because  $A \vdash B$  is in  $H$  when line 11 is to be executed. Then,  $A - B$  must be in  $G$ . Likewise,  $B - C$  must be in  $G$ . However, this contradicts the assumption that  $G$  has a triplex  $(\{A, C\}, B)$ .

**Case 2.2** Assume that the triplex gets removed from  $H$  in lines 13-15. Recall that all the undirected edges in  $H$  at line 16 are in  $G$  by Lemma 3. Therefore, any triplex that gets removed from  $H$  in lines 13-15 cannot exist in  $G$ .

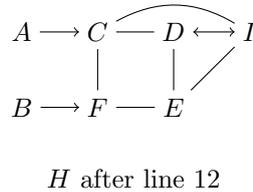
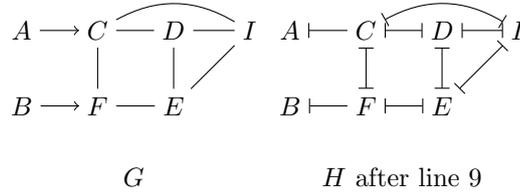
The proofs of the following two lemmas can be found in [10, Lemmas 5 and 6]. The fact that  $G$  is an AMP CG in that work whereas it is an MAMP CG in this work is irrelevant for the proofs. What matters is that both works use the same rules R1-R4.

**Lemma 5.** *After having executed line 9,  $H$  does not have any induced subgraph of the form  $A \overset{\curvearrowright}{\vdash} B \text{ --- } C$ .*

**Lemma 6.** *After having executed line 9, every chordless cycle  $\rho: V_1, \dots, V_n = V_1$  in  $H$  that has an edge  $V_i \vdash V_{i+1}$  also has an edge  $V_j \dashv V_{j+1}$ .*

Lemma 5 is used in the proof of Lemma 6. It is worth noting that one may think that Lemma 5 implies that  $H$  does not have any induced subgraph of the form  $A \overset{\curvearrowright}{\leftrightarrow} B \text{ --- } C$  after having executed line 12 and, thus, that lines 13-15 are not needed. However, this is wrong as the following example illustrates.

*Example 2.* The MAMP CG  $G$  below shows that lines 13-15 are necessary.



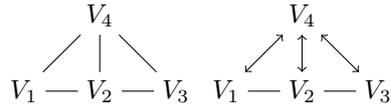
We can now prove the correctness of our algorithm.

**Theorem 1.** *At line 16,  $H$  is an MAMP CG that is triplex equivalent to  $G$ .*

*Proof.* Lemma 1 implies that  $H$  at line 16 has the same adjacencies as  $G$ . Lemma 4 implies that  $H$  at line 16 has the same triplexes as  $G$ . Lemma 6 implies that  $H$  has no semidirected chordless cycle after having executed line 11. This implies that  $H$  has no semidirected chordless cycle at line 16, because all lines 12-15 do is replacing bidirected edges in  $H$  with undirected edges. To see that this in turn implies that  $H$  has no semidirected cycle at line 16, assume to the contrary that  $H$  has no semidirected chordless cycle but it has a semidirected cycle  $\rho : V_1, \dots, V_n = V_1$  with a chord between  $V_i$  and  $V_j$  with  $i < j$ . Then, divide  $\rho$  into the cycles  $\rho_L : V_1, \dots, V_i, V_j, \dots, V_n = V_1$  and  $\rho_R : V_i, \dots, V_j, V_i$ . Note that  $\rho_L$  or  $\rho_R$  is a semidirected cycle. Then,  $H$  has a semidirected cycle that is shorter than  $\rho$ . By repeated application of this reasoning, we can conclude that  $H$  has a semidirected chordless cycle, which is a contradiction. Therefore,  $H$  at line 16 satisfies the constraint C1.

We now show that  $H$  at line 16 satisfies the constraint C2. Assume to the contrary that  $H$  has a cycle  $\rho : V_1, \dots, V_n = V_1$  such that  $V_1 \leftrightarrow V_2$  is in  $H$  and  $V_i - V_{i+1}$  is in  $H$  for all  $1 < i < n$ . Note that  $\rho$  must be of length greater than three by lines 13-15, i.e.  $n > 3$ . Note also that  $V_i - V_{i+1}$  must be in  $G$  for all  $1 < i < n$  by Lemma 3, which implies that  $V_1 - V_2$  is also in  $G$  by the constraints C1 and C2. This implies that  $V_1$  and  $V_3$  are adjacent in  $G$  because, otherwise,  $G$  and  $H$  have not the same triplexes, which contradicts Lemma 4. Then,  $V_1$  and  $V_3$  are adjacent in  $H$  by Lemma 1. In fact,  $V_1 \leftrightarrow V_3$  must be in  $H$  because, otherwise,  $H$  has a cycle of length three that violates the constraint C1 or C2 which, as shown above, is a contradiction. Then,  $H$  has a cycle that violates the constraint C2 and that is shorter than  $\rho$ , namely  $V_1, V_3, \dots, V_n = V_1$ . By repeated application of this reasoning, we can conclude that  $H$  has a cycle of length three that violates the constraint C2 which, as shown above, is a contradiction.

We finally show that  $H$  at line 16 satisfies the constraint C3. Assume to the contrary that  $H$  at line 16 has a subgraph of the form  $V_1 - V_2 - V_3$ , and  $V_2 \leftrightarrow V_4$  is in  $H$  but  $V_1 - V_3$  is not in  $H$ . We show below that  $G$  (respectively  $H$  at line 16) has the graph to the left (respectively right) below as an induced subgraph.



That  $V_1 - V_2 - V_3$  is in  $H$  at line 16 but  $V_1 - V_3$  is not implies that  $V_1$  and  $V_3$  cannot be adjacent in  $H$  because, otherwise,  $H$  violates the constraint C1 or C2 which, as shown above, is a contradiction. This implies that  $V_1$  and  $V_3$  are not adjacent in  $G$  either by Lemma 1. That  $V_1 - V_2 - V_3$  is in  $H$  at line 16 implies that  $V_1 - V_2 - V_3$  is also in  $G$  by Lemma 3. That  $V_2 \leftrightarrow V_4$  is in  $H$  at line 16 implies that  $V_2 \dashv V_4$  is in  $H$  after having executed line 9, which implies that  $V_2 - V_4$  or  $V_2 \leftrightarrow V_4$  is in  $G$  by Lemmas 1 and 2. In fact,  $V_2 - V_4$  must be in  $G$  because, otherwise,  $G$  violates the constraint C3 since, as shown above,  $V_1 - V_2 - V_3$  is in  $G$  but  $V_1 - V_3$  is not. Finally, note that  $V_1$  and  $V_4$  as well as  $V_3$  and  $V_4$  must be adjacent in  $G$  and  $H$  because, otherwise,  $H$  at line 16 does not have the same triplexes as  $G$ , which contradicts Lemma 4. Specifically,  $V_1 - V_4 - V_3$  must be

in  $G$  and  $V_1 \leftrightarrow V_4 \leftrightarrow V_3$  must be in  $H$  at line 16 because, otherwise,  $G$  or  $H$  violates the constraint C1 or C2 which, as shown above, is a contradiction.

However, that  $G$  (respectively  $H$  at line 16) has the graph to the left (respectively right) above as an induced subgraph implies that  $H$  has a triplex  $(\{V_1, V_3\}, V_4)$  that  $G$  has not, which contradicts Lemma 4. Then,  $V_1$  and  $V_3$  must be adjacent in  $H$  which, as shown above, is a contradiction.

## 5 Discussion

MAMP CGs are a recently introduced family of models that is based on graphs that may have undirected, directed and bidirected edges. They unify and generalize AMP and MVR CGs. In this paper, we have presented an algorithm for learning an MAMP CG from a probability distribution  $p$  which is faithful to it. In practice, 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 one that is consistent so that the resulting algorithm is asymptotically correct. We are currently working in the implementation and empirical evaluation of our algorithm. It is worth mentioning that, whereas R1, R2 and R4 only involve three or four nodes, R3 may involve more. Unfortunately, we have not succeeded so far in proving the correctness of our algorithm with a simpler R3. Note that the output of our algorithm would be the same. The only benefit might be a decrease in running time.

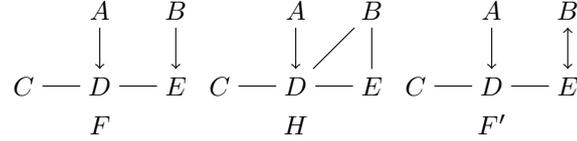
The correctness of our algorithm relies upon the assumption that  $p$  is faithful to some MAMP CG. This is a strong requirement that we would like to weaken, e.g. by replacing it with the milder assumption that  $p$  satisfies the composition property. Specifically,  $p$  satisfies the composition property when  $X \perp_p Y | Z \wedge X \perp_p W | Z \Rightarrow X \perp_p Y \cup W | Z$  for all  $X, Y, Z$  and  $W$  pairwise disjoint subsets of  $V$ . Note that if  $p$  is a Gaussian distribution, then it satisfies the composition property regardless of whether it is faithful or not to some MAMP CG [20, Corollary 2.4].

When making the faithfulness assumption is not reasonable, the correctness of a learning algorithm may be redefined as follows. Given an MAMP CG  $G$ , 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  $Z$  pairwise disjoint subsets of  $V$ . We say that a learning algorithm is correct when it returns an MAMP CG  $H$  such that  $p$  is Markovian with respect to  $H$  and  $p$  is not Markovian with respect to any MAMP CG  $F$  such that  $I(H) \subseteq I(F)$ .

Correct algorithms for learning Bayesian networks and LWF CGs under the composition property assumption exist [3, 9, 13]. 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 with respect to  $H$ , and a second phase that removes single edges from  $H$  until  $p$  is Markovian with respect to  $H$  and  $p$  is not Markovian with respect to any

graph  $F$  such that  $I(H) \subseteq I(F)$ . 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 [8]. Specifically, given two directed and acyclic graphs  $F$  and  $H$  such that  $I(H) \subseteq I(F)$ , Meek's conjecture states that we can transform  $F$  into  $H$  by a sequence of operations such that, 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 [2, Theorem 4]. The extension of Meek's conjecture to LWF CGs was proven to be true in [13, Theorem 1]. The extension of Meek's conjecture to AMP and MVR CGs was proven to be false in [10, Example 1] and [12], respectively. Unfortunately, the extension of Meek's conjecture to MAMP CGs does not hold either, as the following example illustrates.

*Example 3.* The MAMP CGs  $F$  and  $H$  below show that the extension of Meek's conjecture to MAMP CGs does not hold.



We can describe  $I(F)$  and  $I(H)$  by listing all the separators between any pair of distinct nodes. We indicate whether the separators correspond to  $F$  or  $H$  with a superscript. Specifically,

- $\mathcal{S}_{AD}^F = \mathcal{S}_{BE}^F = \mathcal{S}_{CD}^F = \mathcal{S}_{DE}^F = \emptyset$ ,
- $\mathcal{S}_{AB}^F = \{\emptyset, \{C\}, \{D\}, \{E\}, \{C, D\}, \{C, E\}\}$ ,
- $\mathcal{S}_{AC}^F = \{\emptyset, \{B\}, \{E\}, \{B, E\}\}$ ,
- $\mathcal{S}_{AE}^F = \{\emptyset, \{B\}, \{C\}, \{B, C\}\}$ ,
- $\mathcal{S}_{BC}^F = \{\emptyset, \{A\}, \{D\}, \{A, D\}, \{A, D, E\}\}$ ,
- $\mathcal{S}_{BD}^F = \{\emptyset, \{A\}, \{C\}, \{A, C\}\}$ , and
- $\mathcal{S}_{CE}^F = \{\{A, D\}, \{A, B, D\}\}$ .

Likewise,

- $\mathcal{S}_{AD}^H = \mathcal{S}_{BD}^H = \mathcal{S}_{BE}^H = \mathcal{S}_{CD}^H = \mathcal{S}_{DE}^H = \emptyset$ ,
- $\mathcal{S}_{AB}^H = \{\emptyset, \{C\}, \{E\}, \{C, E\}\}$ ,
- $\mathcal{S}_{AC}^H = \{\emptyset, \{B\}, \{E\}, \{B, E\}\}$ ,
- $\mathcal{S}_{AE}^H = \{\emptyset, \{B\}, \{C\}, \{B, C\}\}$ ,
- $\mathcal{S}_{BC}^H = \{\{A, D\}, \{A, D, E\}\}$ , and
- $\mathcal{S}_{CE}^H = \{\{A, D\}, \{A, B, D\}\}$ .

Then,  $I(H) \subseteq I(F)$  because  $\mathcal{S}_{XY}^H \subseteq \mathcal{S}_{XY}^F$  for all  $X, Y \in \{A, B, C, D, E\}$  with  $X \neq Y$ . Moreover, the MAMP CG  $F'$  above is the only MAMP CG that is triplex equivalent to  $F$ , whereas there is no MAMP CG that is triplex equivalent to  $H$ . Obviously, one cannot transform  $F$  or  $F'$  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.** We thank the Reviewers for their thorough reviews. This work is funded by the Center for Industrial Information Technology (CENIIT) and a so-called career contract at Linköping University, and by the Swedish Research Council (ref. 2010-4808).

## References

1. Andersson, S. A., Madigan, D. and Perlman, M. D. Alternative Markov Properties for Chain Graphs. *Scandinavian Journal of Statistics*, 28:33-85, 2001.
2. Chickering, D. M. Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research*, 3:507-554, 2002.
3. Chickering, D. M. and Meek, C. Finding Optimal Bayesian Networks. In *Proceedings of 18th Conference on Uncertainty in Artificial Intelligence*, 94-102, 2002.
4. Cox, D. R. and Wermuth, N. *Multivariate Dependencies - Models, Analysis and Interpretation*. Chapman & Hall, 1996.
5. Koster, J. T. A. Marginalizing and Conditioning in Graphical Models. *Bernoulli*, 8:817-840, 2002.
6. Lauritzen, S. L. *Graphical Models*. Oxford University Press, 1996.
7. Ma, Z., Xie, X. and Geng, Z. Structural Learning of Chain Graphs via Decomposition. *Journal of Machine Learning Research*, 9:2847-2880, 2008.
8. Meek, C. *Graphical Models: Selecting Causal and Statistical Models*. PhD thesis, Carnegie Mellon University, 1997.
9. 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.
10. Peña, J. M. Learning AMP Chain Graphs and some Marginal Models Thereof under Faithfulness. *International Journal of Approximate Reasoning*, 55:1011-1021, 2014.
11. Peña, J. M. Marginal AMP Chain Graphs. *International Journal of Approximate Reasoning*, 55:1185-1206, 2014.
12. Peña, J. M. Learning Multivariate Regression Chain Graphs under Faithfulness: Addendum. Available at the author's website, 2014.
13. Peña, J. M., Sonntag, D. and Nielsen, J. D. An Inclusion Optimal Algorithm for Chain Graph Structure Learning. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, 778-786, 2014.
14. Richardson, T. and Spirtes, P. Ancestral Graph Markov Models. *The Annals of Statistics*, 30:962-1030, 2002.
15. Sadeghi, K. and Lauritzen, S. L. Markov Properties for Mixed Graphs. *Bernoulli*, 20:676-696, 2014.
16. Sonntag, D. and Peña, J. M. Learning Multivariate Regression Chain Graphs under Faithfulness. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, 299-306, 2012.

17. Sonntag, D. and Peña, J. M. Chain Graph Interpretations and their Relations. In *Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty*, 510-521, 2013.
18. Spirtes, P., Glymour, C. and Scheines, R. *Causation, Prediction, and Search*. Springer-Verlag, 1993.
19. Studený, M. A Recovery Algorithm for Chain Graphs. *International Journal of Approximate Reasoning*, 17:265-293, 1997a.
20. Studený, M. *Probabilistic Conditional Independence Structures*. Springer, 2005.