

# Learning AMP chain graphs and some marginal models thereof under faithfulness

Jose M Pena

**Linköping University Post Print**



N.B.: When citing this work, cite the original article.

Original Publication:

Jose M Pena, Learning AMP chain graphs and some marginal models thereof under faithfulness, 2014, International Journal of Approximate Reasoning, (55), 4, 1011-1021.

<http://dx.doi.org/10.1016/j.ijar.2014.01.003>

Copyright: Elsevier

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-106668>

# LEARNING AMP CHAIN GRAPHS AND SOME MARGINAL MODELS THEREOF 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 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. Moreover, we 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.

We also study the problem of how to represent the result of marginalizing out some nodes in an AMP CG. We introduce a new family of graphical models that solves this problem partially. We name this new family maximal covariance-concentration graphs because it includes both covariance and concentration graphs as subfamilies.

## 1. INTRODUCTION

This paper deals with chain graphs (CGs) under the Andersson-Madigan-Perlman (AMP) interpretation (Andersson et al., 2001). Two other interpretations exist in the literature, namely the Lauritzen-Wermuth-Frydenberg (LWF) interpretation (Lauritzen, 1996) and the multivariate regression (MVR) interpretation (Cox and Wermuth, 1996). The AMP and LWF interpretations 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). Likewise, neither the AMP interpretation subsumes the MVR interpretation nor vice versa (Sonntag and Peña, 2013, Theorems 4 and 5).

In this paper, 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, algorithms for learning LWF CGs under faithfulness already exist (Ma et al., 2008; Studený, 1997a). In fact, we have recently developed an algorithm for learning LWF CGs under the milder composition property assumption (Peña et al., 2012). We have also recently developed an algorithm for learning MVR CGs under the faithfulness assumption (Sonntag and Peña, 2012).

As Richardson and Spirtes (2002, Section 9.4) show, a desirable feature that AMP CGs lack is that of being closed under marginalization (a.k.a the precollapsibility property (Studený, 1997b)). That is, the independence model resulting from marginalizing out some nodes in an AMP CG may not be represented by any other AMP CG. This leads us to the problem of how to represent the result of marginalizing out some nodes in an AMP CG. Of course, one may decide to continue working with the AMP CG and treat the marginalized nodes as latent nodes. This solution relies upon one having access to the AMP CG. Thus, it does not solve the problem if one knows that there is an underlying AMP CG but does not have access to it. As far as we know, this problem has been studied for directed and acyclic graphs by Richardson and Spirtes (2002) but not for AMP CGs. In this paper, we present the

partial solution to this problem that we have obtained so far. Specifically, we introduce and study a new family of graphical models that we call maximal covariance-concentration graphs (MCCGs). MCCGs solve the problem at hand partially, because each of them represents the result of marginalizing out some nodes in some AMP CG. Unfortunately, MCCGs do not solve the problem completely, because they do not represent the result of marginalizing out any nodes in any AMP CG.

MCCGs consist of undirected and bidirected edges, and they unify and generalize covariance and concentration graphs, hence the name. Concentration graphs (a.k.a Markov networks) were introduced by Pearl (1988) to represent independence models. Specifically, the concentration graph of a probability distribution  $p$  is the undirected graph  $G$  where two nodes are not adjacent if and only if their corresponding random variables are independent in  $p$  given the rest of the random variables. Graphical criteria for reading dependencies and independencies from  $G$  (under certain assumptions about  $p$ ) have been proposed (Bouckaert, 1995; Pearl, 1988; Peña et al., 2009). Likewise, covariance graphs (a.k.a bidirected graphs) were introduced by Cox and Wermuth (1996) to represent independence models. Specifically, the covariance graph of a probability distribution  $p$  is the bidirected graph  $G$  where two nodes are not adjacent if and only if their corresponding random variables are marginally independent in  $p$ . Graphical criteria for reading dependencies and independencies from  $G$  (under certain assumptions about  $p$ ) have been proposed (Banerjee and Richardson, 2003; Kauermann, 1996; Peña, 2013a).

If we focus on Gaussian probability distributions, then one could say that the covariance graph of a Gaussian probability distribution models its covariance matrix, whereas its concentration graph models its concentration matrix. We think that Gaussian probability distributions would be modeled more accurately if their covariance and concentration matrices were modeled jointly by a single graph. This is something one can do with MCCGs.

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

## 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 simple, i.e. they contain at most one edge between any pair of nodes. 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, 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  $G$ . 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 spouses of a set of nodes  $X$  of  $G$  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 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  $V_1, \dots, V_n$  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 a cycle if  $V_n = V_1$ . A cycle has a chord if two non-consecutive nodes of the cycle are adjacent in  $G$ . 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  $V_i \rightarrow V_{i+1}$  is in  $G$  for some  $1 \leq i < n$ . A chain graph (CG) is a graph whose every edge is undirected or directed, and that has no semidirected cycles. A set

of nodes of a graph is complete if there is an undirected edge between every pair of nodes in the set. A set of nodes of a graph is undirectly (respectively bidirectly) connected if there exists a route in the graph between every pair of nodes in the set such that all the edges in the route are undirected (respectively bidirected). An undirected (respectively bidirected) connectivity component of a graph is an undirectly (respectively bidirectly) connected set that is maximal (with respect to set inclusion). The undirected connectivity component a node  $A$  of a graph  $G$  belongs to is denoted as  $co_G(A)$ . The subgraph of  $G$  induced by a set of its nodes  $X$ , denoted as  $G_X$ , is the graph over  $X$  that has all and only the edges in  $G$  whose both ends are in  $X$ . An immorality in a CG is an induced subgraph of the form  $A \rightarrow B \leftarrow C$ . A flag in a CG is an induced subgraph of the form  $A \rightarrow B - C$ . If a CG  $G$  has an induced subgraph of the form  $A \rightarrow B \leftarrow C$ ,  $A \rightarrow B - C$  or  $A - B \leftarrow C$ , then we say that the triplex  $(\{A, C\}, B)$  is in  $G$ . Two CGs are triplex equivalent if and only if they have the same adjacencies and the same triplexes.

Let  $X, Y, Z$  and  $W$  denote four pairwise disjoint subsets of  $V$ . An independence model  $M$  is a set of statements  $X \perp_M Y | Z$ .  $M$  satisfies the graphoid properties if it satisfies the following properties:

- Symmetry  $X \perp_M Y | Z \Rightarrow Y \perp_M X | Z$ .
- Decomposition  $X \perp_M Y \cup W | Z \Rightarrow X \perp_M Y | Z$ .
- Weak union  $X \perp_M Y \cup W | Z \Rightarrow X \perp_M Y | Z \cup W$ .
- Contraction  $X \perp_M Y | Z \cup W \wedge X \perp_M W | Z \Rightarrow X \perp_M Y \cup W | Z$ .
- Intersection  $X \perp_M Y | Z \cup W \wedge X \perp_M W | Z \cup Y \Rightarrow X \perp_M Y \cup W | Z$ .

Two other properties that  $M$  may satisfy are the following:

- Composition  $X \perp_M Y | Z \wedge X \perp_M W | Z \Rightarrow X \perp_M Y \cup W | Z$ .
- Weak transitivity  $X \perp_M Y | Z \wedge X \perp_M Y | Z \cup K \Rightarrow X \perp_M K | Z \vee K \perp_M Y | Z$  with  $K \in V \setminus X \setminus Y \setminus Z$ .

We say that an independence model is a WTC graphoid when it satisfies the seven previous properties. 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 Markovian with respect to an independence model  $M$  when  $X \perp_p Y | Z$  if  $X \perp_M Y | Z$  for all  $X, Y$  and  $Z$  pairwise disjoint subsets of  $V$ . We say that  $p$  is faithful to  $M$  when  $X \perp_p Y | Z$  if and only if  $X \perp_M Y | Z$  for all  $X, Y$  and  $Z$  pairwise disjoint subsets of  $V$ . Any probability distribution  $p$  satisfies the first four previous properties. If  $p$  is faithful to a CG, then it also satisfies the last three previous properties.<sup>1</sup>

A node  $B$  in a route  $\rho$  in a CG is called a head-no-tail node in  $\rho$  if  $A \rightarrow B \leftarrow C$ ,  $A \rightarrow B - C$ , or  $A - B \leftarrow C$  is a subroute of  $\rho$  (note that maybe  $A = C$  in the first case). A node  $B$  in  $\rho$  is called a non-head-no-tail node in  $\rho$  if  $A \leftarrow B \rightarrow C$ ,  $A \leftarrow B \leftarrow C$ ,  $A \leftarrow B - C$ ,  $A \rightarrow B \rightarrow C$ ,  $A - B \rightarrow C$ , or  $A - B - C$  is a subroute of  $\rho$  (note that maybe  $A = C$  in the first and last cases). Note that to classify  $B$  as a (non-)head-no-tail node in  $\rho$ , one has to consider the edge ends at  $B$  as well as at  $A$  and  $C$ . Note also that  $B$  may be both a head-no-tail and a non-head-no-tail node in  $\rho$ , e.g. take  $\rho$  to be  $A \rightarrow B \leftarrow C \rightarrow B \rightarrow D$ . Let  $X, Y$  and  $Z$  denote three pairwise 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 non-head-no-tail node in  $\rho$  is not in  $Z$ .<sup>2</sup> 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  $X \perp_G Y | Z$ .<sup>3</sup> We denote by  $X \not\perp_G Y | Z$  that  $X \perp_G Y | Z$  does not hold. The independence model induced by  $G$ , denoted as  $I(G)$ , is

<sup>1</sup>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, composition and weak transitivity properties (Studeny, 2005, Proposition 2.1 and Corollaries 2.4 and 2.5).

<sup>2</sup>Note that if a node is both a head-no-tail and a non-head-no-tail node in  $\rho$ , then  $\rho$  is not  $Z$ -open.

<sup>3</sup>See (Andersson et al., 2001, Remark 3.1) for the equivalence of this and the standard definition of separation.

TABLE 1. Algorithm for learning AMP CGs.

---

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  $|(ad_H(A) \cup ad_H(ad_H(A))) \setminus B| \geq l$
- 5     If there is some  $S \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \setminus B$  such that  $|S| = l$  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  $A \dashv B$  (respectively  $A \dashv B$ ) in  $H$  with  $A \rightarrow B$  (respectively  $A \leftarrow B$ )

---

the set of separation statements  $X \perp_G Y | Z$ . If two CGs  $G$  and  $H$  are triplex equivalent, then  $I(G) = I(H)$ .<sup>4</sup>

### 3. ALGORITHM FOR LEARNING AMP CGS

In this section, we present an algorithm for learning an AMP CG a given probability distribution is faithful to. The 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 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, 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  $\dashv$  or  $\dashv$ , 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 rule are applied. Note that the ends of some of the edges in the rules are labeled with a circle such as in  $\dashv\circ$  or  $\circ\dashv$ . 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.

**3.1. 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.

---

<sup>4</sup>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 with respect to  $H$  and  $G$ , respectively, by Andersson et al. (2001, Theorem 5) and Levitz et al. (2001, Theorem 4.1).

TABLE 2. Rules R1-R4 in the algorithm for learning AMP 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:	
R4:	 $\wedge A \in S_{CD}$

**Lemma 1.** *Let  $p$  denote a probability distribution that satisfies the intersection and composition properties. Then,  $p$  is Markovian with respect to a CG  $G$  if and only if  $p$  satisfies the following conditions:*

- C1:  $A \perp_p co_G(A) \setminus A \setminus ne_G(A) | pa_G(A \cup ne_G(A)) \cup ne_G(A)$  for all  $A \in V$ , and  
C2:  $A \perp_p V \setminus A \setminus de_G(A) \setminus pa_G(A) | pa_G(A)$  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 with respect to  $G$  if and only if  $p$  satisfies the following conditions:

- L1:  $A \perp_p co_G(A) \setminus A \setminus ne_G(A) | [V \setminus co_G(A) \setminus de_G(co_G(A))] \cup ne_G(A)$  for all  $A \in V$ , and  
L2:  $A \perp_p V \setminus co_G(A) \setminus de_G(co_G(A)) \setminus pa_G(A) | pa_G(A)$  for all  $A \in V$ .

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

1.  $B \perp_p V \setminus co_G(B) \setminus de_G(co_G(B)) \setminus pa_G(B) | pa_G(B)$  for all  $B \in A \cup ne_G(A)$  by L2.
2.  $B \perp_p V \setminus co_G(B) \setminus de_G(co_G(B)) \setminus pa_G(A \cup ne_G(A)) | pa_G(A \cup ne_G(A))$  for all  $B \in A \cup ne_G(A)$  by weak union on 1.
3.  $A \cup ne_G(A) \perp_p V \setminus co_G(A) \setminus de_G(co_G(A)) \setminus pa_G(A \cup ne_G(A)) | pa_G(A \cup ne_G(A))$  by repeated application of symmetry and composition on 2.
4.  $A \perp_p V \setminus co_G(A) \setminus de_G(co_G(A)) \setminus pa_G(A \cup ne_G(A)) | pa_G(A \cup ne_G(A)) \cup ne_G(A)$  by symmetry and weak union on 3.
5.  $A \perp_p co_G(A) \setminus A \setminus ne_G(A) | [V \setminus co_G(A) \setminus de_G(co_G(A))] \cup ne_G(A)$  by L1.
6.  $A \perp_p [co_G(A) \setminus A \setminus ne_G(A)] \cup [V \setminus co_G(A) \setminus de_G(co_G(A)) \setminus pa_G(A \cup ne_G(A))] | pa_G(A \cup ne_G(A)) \cup ne_G(A)$  by contraction on 4 and 5.
7.  $A \perp_p co_G(A) \setminus A \setminus ne_G(A) | pa_G(A \cup ne_G(A)) \cup ne_G(A)$  by decomposition on 6.

We now prove the only if part.

8.  $A \perp_p co_G(A) \setminus A \setminus ne_G(A) | pa_G(A \cup ne_G(A)) \cup ne_G(A)$  by C1.
9.  $A \perp_p [V \setminus co_G(A) \setminus de_G(co_G(A)) \setminus pa_G(A \cup ne_G(A))] \cup [co_G(A) \setminus A \setminus ne_G(A)] | pa_G(A \cup ne_G(A)) \cup ne_G(A)$  by composition on 4 and 8.
10.  $A \perp_p co_G(A) \setminus A \setminus ne_G(A) | [V \setminus co_G(A) \setminus de_G(co_G(A))] \cup ne_G(A)$  by weak union on 9.

□

**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  $S \subseteq V \setminus [A \cup B]$  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,  $A \perp_p co_G(A) \setminus A \setminus ne_G(A)|pa_G(A \cup ne_G(A)) \cup ne_G(A)$  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,  $pa_G(A \cup ne_G(A)) \cup ne_G(A) \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \setminus B$  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,  $A \perp_p V \setminus A \setminus de_G(A) \setminus pa_G(A)|pa_G(A)$  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,  $pa_G(A) \subseteq ad_H(A) \setminus B$  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 in certain sense.

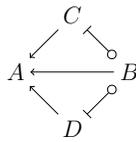
**Lemma 3.** *The rules R1-R4 are sound in the sense that they block only those edge ends that are not arrowheads in  $G$ .*

*Proof.* According to the antecedent of R1,  $G$  has a triplex  $(\{A, C\}, B)$ . Then,  $G$  has an induced subgraph of the form  $A \rightarrow B \leftarrow C$ ,  $A \rightarrow B - C$  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)  $A \rightarrow B$  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 descending route from  $A$  to  $B$ , and (ii)  $A \in ad_G(B)$ . Then,  $A \rightarrow B$  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  $B \rightarrow C$  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, it would have a semidirected cycle. However, this induced subgraph contradicts that  $A \in S_{CD}$ .  $\square$

**Lemma 4.** *After line 10,  $G$  and  $H$  have the same triplexes. Moreover,  $H$  has all the immoralities that are 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 if and only if, when line 10 is executed,  $H$  has an induced subgraph of one of the following forms:

$$A \dashv B \dashv C \quad A \dashv B \dashv C.$$

Note that Lemma 2 implies that  $A$  is adjacent to  $B$  in  $G$ ,  $B$  is adjacent to  $C$  in  $G$ , and that  $A$  is not adjacent to  $C$  in  $G$ . This together with the assumption made above that  $G$  has no triplex  $(\{A, C\}, B)$  implies that  $B \in S_{AC}$  because, otherwise, the route  $A, B, C$  is  $S_{AC}$ -open

in  $G$  contradicting  $A \perp_{GC} | S_{AC}$ . Now, note that the first, second and fifth induced subgraphs above are impossible because, otherwise,  $A \circ \rightarrow B$  would be in  $H$  by R2. Likewise, the third and fourth induced subgraphs above are impossible because, otherwise,  $B \rightarrow C$  would be in  $H$  by R2.

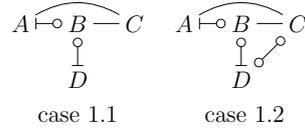
We now prove that any triplex  $(\{A, C\}, B)$  in  $G$  is in  $H$ . Let the triplex be of the form  $A \rightarrow B \leftarrow C$ . Hence,  $B \notin S_{AC}$ . Then, when line 10 is executed,  $A \rightarrow B \leftarrow C$  is in  $H$  by R1, and neither  $A \rightarrow B$  nor  $B \rightarrow C$  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  $A \rightarrow B - C$ . Hence,  $B \notin S_{AC}$ . Then, when line 10 is executed,  $A \rightarrow B \leftarrow C$  is in  $H$  by R1, and  $A \rightarrow B$  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  $H$ .  $\square$

**Lemma 5.** *After line 9,  $H$  does not have any induced subgraph of the form  $\widehat{A \rightarrow B - C}$ .*

*Proof.* Assume to the contrary that the lemma does not hold. We interpret the execution of line 9 as a sequence of block additions and, for the rest of the proof, one particular sequence of these block additions is fixed. Fixing this sequence is a crucial point upon which some important later steps of the proof are based. Since there may be several induced subgraphs of  $H$  of the form under study after line 9, let us consider any of the induced subgraphs

$\widehat{A \rightarrow B - C}$  that appear firstly during execution of line 9 and fix it for the rest of the proof. Now, consider the following cases.

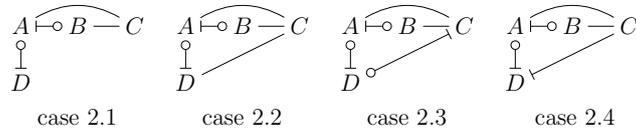
**Case 1:** Assume that  $A \rightarrow B$  is in  $H$  due to R1. Then, after R1 was applied to  $A \circ \rightarrow B$ ,  $H$  had an induced subgraph of one of the following forms:



**Case 1.1:** If  $B \notin S_{CD}$  then  $B \rightarrow C$  is in  $H$  by R1, else  $B \rightarrow C$  is in  $H$  by R2. Either case is a contradiction.

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

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



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

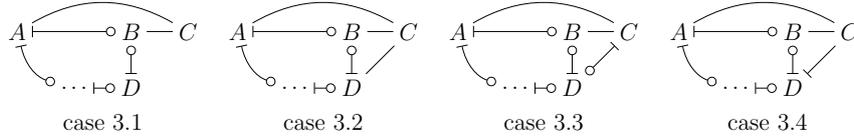
**Case 2.2:** Note that  $\widehat{D \rightarrow A - C}$  cannot be an induced subgraph of  $H$  after line 9 because, otherwise, it would contradict the assumption that  $\widehat{A \rightarrow B - C}$  is one of the firstly induced subgraph of that form that appeared during the execution of line 9. Then,  $A \rightarrow C$ ,  $A \rightarrow C$ ,  $D \circ \rightarrow C$  or  $D \rightarrow C$  must be in  $H$  after line 9. However, either of the first two cases is a contradiction. The third case can be reduced to Case 2.3 as follows. The fourth case can be reduced to Case 2.4 similarly. The third case implies that the block at  $C$  in  $D \circ \rightarrow C$  is added at some moment in the execution of line 9. This moment must happen later than immediately after adding the block at  $A$  in  $A \rightarrow B$ , because immediately after

adding this block the situation is the one depicted by the above figure for Case 2.2. Then, when the block at  $C$  in  $D \circlearrowleft C$  is added, the situation is the one depicted by the above figure for Case 2.3.

**Case 2.3:** Assume that the situation of this case occurs at some moment in the execution of line 9. Then,  $A \rightarrow C$  is in  $H$  after the execution of line 9 by R3, which is a contradiction.

**Case 2.4:** Assume that the situation of this case occurs at some moment in the execution of line 9. If  $C \notin S_{BD}$  then  $B \vdash C$  is in  $H$  after the execution of line 9 by R1, else  $B \rightarrow C$  is in  $H$  after the execution of line 9 by R2. Either case is a contradiction.

**Case 3:** Assume that  $A \vdash B$  is in  $H$  due to R3. Then, after R3 was applied to  $A \circlearrowleft B$ ,  $H$  had a subgraph of one of the following forms, where possible additional edges between  $C$  and internal nodes of the route  $A \vdash \dots \vdash D$  are not shown:



Note that  $C$  cannot belong to the route  $A \vdash \dots \vdash D$  because, otherwise, R3 could not have been applied since the cycle  $A \vdash \dots \vdash D \vdash B \rightarrow A$  would not have been chordless.

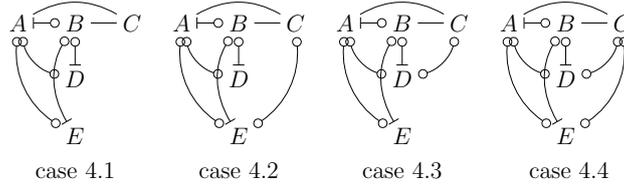
**Case 3.1:** If  $B \notin S_{CD}$  then  $B \rightarrow C$  is in  $H$  by R1, else  $B \vdash C$  is in  $H$  by R2. Either case is a contradiction.

**Case 3.2:** Note that  $D \vdash B \rightarrow C$  cannot be an induced subgraph of  $H$  after line 9 because, otherwise, it would contradict the assumption that  $A \vdash B \rightarrow C$  is one of the firstly induced subgraph of that form that appeared during the execution of line 9. Then,  $B \vdash C$ ,  $B \rightarrow C$ ,  $D \circlearrowleft C$  or  $D \vdash C$  must be in  $H$  after line 9. However, either of the first two cases is a contradiction. The third case can be reduced to Case 3.3 as follows. The fourth case can be reduced to Case 3.4 similarly. The third case implies that the block at  $C$  in  $D \circlearrowleft C$  is added at some moment in the execution of line 9. This moment must happen later than immediately after adding the block at  $A$  in  $A \vdash B$ , because immediately after adding this block the situation is the one depicted by the above figure for Case 3.2. Then, when the block at  $C$  in  $D \circlearrowleft C$  is added, the situation is the one depicted by the above figure for Case 3.3.

**Case 3.3:** Assume that the situation of this case occurs at some moment in the execution of line 9. Then,  $B \rightarrow C$  is in  $H$  after the execution of line 9 by R3, which is a contradiction.

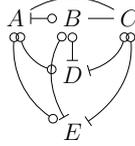
**Case 3.4:** Assume that the situation of this case occurs at some moment in the execution of line 9. Note that  $C$  cannot be adjacent to any node of the route  $A \vdash \dots \vdash D$  besides  $A$  and  $D$ . To see it, assume to the contrary that  $C$  is adjacent to some nodes  $E_1, \dots, E_n \neq A, D$  of the route  $A \vdash \dots \vdash D$ . Assume without loss of generality that  $E_i$  is closer to  $A$  in the route than  $E_{i+1}$  for all  $1 \leq i < n$ . Now, note that  $E_n \vdash C$  must be in  $H$  after the execution of line 9 by R3. This implies that  $E_{n-1} \vdash C$  must be in  $H$  after the execution of line 9 by R3. By repeated application of this argument, we can conclude that  $E_1 \vdash C$  must be in  $H$  after the execution of line 9 and, thus,  $A \vdash C$  must be in  $H$  after the execution of line 9 by R3, which is a contradiction.

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



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

**Case 4.4:** Assume that  $C \in S_{DE}$ . Then,  $B \rightarrow C$  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,  $A \leftarrow C$  is in  $H$  by R4, which is a contradiction. □

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

*Proof.* Assume for a contradiction that  $\rho$  is of the length three such that  $V_1 \leftarrow V_2$  occur and neither  $V_2 \rightarrow V_3$  nor  $V_1 \leftarrow V_3$  occur. Note that  $V_2 \rightarrow V_3$  cannot occur either because, otherwise,  $V_1 \leftarrow V_3$  or  $V_1 \rightarrow V_3$  must occur by R3. Since the former contradicts the assumption, then the latter must occur. However, this implies that  $V_1 \rightarrow V_2$  must occur by R3, which contradicts the assumption. Similarly,  $V_1 \rightarrow V_3$  cannot occur either. Then,  $\rho$  is of one of the following forms:

$$V_1 \leftarrow V_2 \rightarrow V_3 \quad V_1 \leftarrow V_2 \rightarrow V_3 \quad V_1 \leftarrow V_2 \leftarrow V_3.$$

The first form is impossible by Lemma 5. The second form is impossible because, otherwise,  $V_2 \rightarrow V_3$  would occur by R3. The third form is impossible because, otherwise,  $V_1 \leftarrow V_3$  would occur by R3. Thus, the lemma holds for cycles of length three.

Assume for a contradiction that  $\rho$  is of length greater than three and has an edge  $V_i \leftarrow V_{i+1}$  but no edge  $V_j \rightarrow V_{j+1}$ . Note that if  $V_l \leftarrow V_{l+1} \rightarrow V_{l+2}$  is a subroute of  $\rho$ , then either  $V_{l+1} \leftarrow V_{l+2}$  or  $V_{l+1} \rightarrow V_{l+2}$  is in  $\rho$  by R1 and R2. Since  $\rho$  has no edge  $V_j \rightarrow V_{j+1}$ ,  $V_{l+1} \rightarrow V_{l+2}$  is in  $\rho$ . By repeated application of this reasoning together with the fact that  $\rho$  has an edge  $V_i \leftarrow V_{i+1}$ , we can conclude that every edge in  $\rho$  is  $V_k \leftarrow V_{k+1}$ . Then, by repeated application of R3, observe that every edge in  $\rho$  is  $V_k \rightarrow V_{k+1}$ , which contradicts the assumption. □

**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 chordless cycle, which implies that  $H$  has no semidirected cycle. To see the latter implication, assume to the contrary that  $H$  has no semidirected chordless cycle but that 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. □

## 4. MAXIMAL COVARIANCE-CONCENTRATION GRAPHS

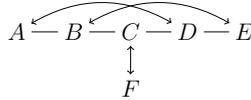
As mentioned in the introduction, AMP CGs are not closed under marginalization, which leads us to the problem of how to represent the result of marginalizing out some nodes in an AMP CG. In this section, we present the partial solution to this problem that we have obtained so far. Specifically, we introduce and study a new family of graphical models that we call maximal covariance-concentration graphs. These new models solve the problem at hand partially, because each of them represents the result of marginalizing out some nodes in some AMP CG. Unfortunately, our new models do not solve the problem completely, because they do not represent the result of marginalizing out any nodes in any AMP CG. An extended version of this section can be found in (Peña, 2013b).

First, we define covariance-concentration graphs (CCGs) as graphs whose every edge is undirected or bidirected. A node  $B$  in a path  $\rho$  in a CCG is called a triplex node in  $\rho$  if  $A \leftrightarrow B \leftrightarrow C$ ,  $A \leftrightarrow B - C$  or  $A - B \leftrightarrow C$  is a subpath of  $\rho$ . Let  $X$ ,  $Y$  and  $Z$  denote three pairwise disjoint subsets of  $V$ . A path  $\rho$  in a CCG  $G$  is said to be  $Z$ -open when

- every triplex node in  $\rho$  is in  $Z$ , and
- every non-triplex node in  $\rho$  is not in  $Z$  or has some spouse in  $G$ .

When there is no path 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  $Z$  and denote it as  $X \perp_G Y | Z$ . We denote by  $X \not\perp_G Y | Z$  that  $X \perp_G Y | Z$  does not hold. The independence model induced by  $G$  is the set of separations  $X \perp_G Y | Z$ .

Typically, every missing edge in a graphical model corresponds to a separation. However, this is not true for CCGs. For instance, the CCG  $G$  below does not contain any edge between  $B$  and  $D$  but  $B \not\perp_G D | Z$  for all  $Z \subseteq V \setminus \{B, D\}$ . Likewise,  $G$  does not contain any edge between  $A$  and  $E$  but  $A \not\perp_G E | Z$  for all  $Z \subseteq V \setminus \{A, E\}$ .



In order to avoid the problem above, we focus in this paper on what we call maximal CCGs (MCCGs), which are those CCGs that have

- no induced subgraph  $A - C - B$  such that  $C$  has some spouse, and
- no cycle  $A - \dots - B \leftrightarrow A$ .

Hereinafter, we refer to the two constraints on CCGs above as C1 and C2, respectively. Now, every missing edge in a MCCG corresponds to a separation (Peña, 2013b, Theorem 3). So, no edge can be added to a MCCG without changing the independence model induced by it, hence the name. Note that a MCCG  $G$  represents the same separations over  $V$  as the AMP CG  $H$  obtained by replacing every bidirected edge  $A \leftrightarrow B$  in  $G$  with  $A \leftarrow H_{AB} \rightarrow B$ . Therefore,  $G$  represents the marginal independence model of  $H$  over  $V$ . Note also that both covariance and concentration graphs are MCCGs, and the definitions of separation for covariance and concentration graphs are special cases of the one introduced above for MCCGs (recall Section 1). Therefore, MCCGs unify and generalize covariance and concentration graphs.

Note that if a MCCG has a subgraph  $A - C - B$  such that  $C$  has some spouse, then the constraint C1 implies that there must be an edge between  $A$  and  $B$  in the MCCG, whereas the constraint C2 implies that the edge must be undirected. Therefore, if a MCCG has a path  $A = V_1 - V_2 - \dots - V_n = B$  such that  $V_i$  has some spouse for all  $1 < i < n$ , then the edge  $V_1 - V_n$  must be in the MCCG. Therefore, the independence model induced by a MCCG is the same whether we use the definition of  $Z$ -open path above or the following simpler one. A path  $\rho$  in a MCCG is said to be  $Z$ -open when

- every triplex node in  $\rho$  is in  $Z$ , and

- every non-triplex node in  $\rho$  is not in  $Z$ .

Finally, it is worth mentioning that we show in (Peña, 2013b) that MCCGs enjoy the following interesting properties.

- We show that the independence models induced by MCCGs are not arbitrary in the probabilistic framework because, for any MCCG  $G$ , there exists a regular Gaussian probability distribution that is faithful to  $G$ .
- We show that the independence models induced by MCCGs are WTC graphoids.
- We show that MCCGs are closed under marginalization.
- In addition to the global Markov property introduced above, we also introduce local and pairwise Markov properties for MCCGs. Moreover, we prove that the three properties are equivalent in certain sense.
- We characterize when two MCCGs are Markov equivalent, and show that every Markov equivalence class of MCCGs has a distinguished member which can easily be obtained from any other member in the class.
- We present a constraint based algorithm for learning a MCCG a given probability distribution is faithful to. The algorithm actually returns the distinguished member of the Markov equivalence class of the MCCG in the input.
- We present a graphical criterion for reading dependencies from a MCCG of a probability distribution that is a WTC graphoid, e.g. a Gaussian probability distribution. We prove that the criterion is sound and complete in certain sense.
- We assess the merits of MCCGs with respect to other families of graphical models such as covariance graphs, concentration graphs, maximal ancestral graphs (Richardson and Spirtes, 2002), summary graphs (Cox and Wermuth, 1996) and MC graphs (Koster, 2002). We reach the following conclusions.
  - As mentioned above, MCCGs unify and generalize covariance and concentration graphs. We show that there are cases where a MCCG of a probability distribution  $p$  can identify more (in)dependencies in  $p$  than the covariance graph and the concentration graph of  $p$  jointly.
  - We show that MCCGs are a subfamily of maximal ancestral graphs. However, we also show that there are cases where a MCCG is a more natural representation of the domain at hand than a maximal ancestral graph and, thus, MCCGs are still worth consideration.
  - Maximal ancestral graphs are a subfamily of summary graphs and MC graphs and, thus, so are MCCGs. However, the last two families have a rather counterintuitive and undesirable feature: Not every missing edge corresponds to a separation (Richardson and Spirtes, 2002, p. 1023). We show that MCCGs do not have this disadvantage.

## 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

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

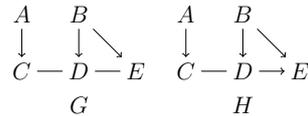
Unfortunately, we have not succeeded so far in proving the correctness of our algorithm with such a simpler rule. Note that the output of our algorithm will be the same whether we keep R3 or we replace it with a simpler sound rule. The only benefit of the simpler rule may be a decrease in running time.

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  $G$ .*

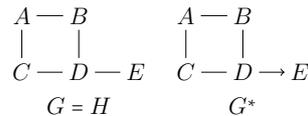
*Proof.* Note that every flag  $A \rightarrow B - C$  in  $H$  after line 10 is due to an induced subgraph of  $H$  of the form  $A \leftarrow B \mapsto C$  after line 9 because  $A \leftarrow B - C$  is excluded by R1 and R2. Note also that all the blocks in  $H$  follow from the adjacencies and triplexes 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.  $\square$

A CG whose every flag is in every other triplex equivalent CG is called a deflagged graph by Roverato and Studený (2006, Proposition 8). Therefore, the lemma above implies that our algorithm outputs a deflagged graph. Note that there may be several deflagged graphs that are triplex equivalent to  $G$ . Unfortunately, not every directed edge in the output of our algorithm is in every deflagged graph that is triplex equivalent to  $G$ , as the following example illustrates (note that both  $G$  and  $H$  are deflagged graphs).



Therefore, our algorithm outputs a deflagged graph but not what Roverato and Studený (2006) call 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).

Another distinguished member of a class of triplex equivalent CGs is the so-called essential graph  $G^*$  (Andersson and Perlman, 2006): An edge  $A \rightarrow B$  is in  $G^*$  if and only if  $A \leftarrow B$  is in no member of the class. Unfortunately, our algorithm does not output an essential graph either, as the following example illustrates.

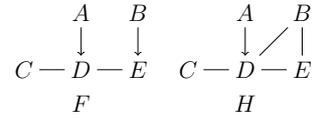


It is worth mentioning that a characterization of essential graphs that is more efficient than the one introduced above is available (Andersson and Perlman, 2006, Theorem 5.1). Also, an efficient algorithm for constructing the essential graph from any member of the class has been proposed (Andersson and Perlman, 2004, Section 7). As far as we know, the correctness of the algorithm has not been proven though.

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 section (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 CG  $F$  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 (Meek, 1997). 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 (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.

**Example 1.** Consider the AMP CGs  $F$  and  $H$  below.



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  $X \neq Y$ . 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

We would like to thank the anonymous Reviewers and specially Reviewer 1 for their comments. This work is funded by the Center for Industrial Information Technology (CENIIT) and a so-called career contract at Linköping University, by the Swedish Research Council (ref. 2010-4808), and by FEDER funds and the Spanish Government (MICINN) through the project TIN2010-20900-C04-03.

## REFERENCES

- Andersson, S. A., Madigan, D. and Perlman, M. D. Alternative Markov Properties for Chain Graphs. *Scandinavian Journal of Statistics*, 28:33-85, 2001.
- Andersson, S. A. and Perlman, M. D. Characterizing Markov Equivalent Classes for AMP Chain Graph Models. Technical Report 453, University of Washington, 2004. Available at <http://www.stat.washington.edu/www/research/reports/2004/tr453.pdf>.
- Andersson, S. A. and Perlman, M. D. Characterizing Markov Equivalent Classes for AMP Chain Graph Models. *The Annals of Statistics*, 34:939-972, 2006.
- Banerjee, M. and Richardson, T. On a Dualization of Graphical Gaussian Models: A Correction Note. *Scandinavian Journal of Statistics*, 30:817-820, 2003.
- Bouckaert, R. R. *Bayesian Belief Networks: From Construction to Inference*. PhD Thesis, University of Utrecht, 1995.
- 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.
- Cox, D. R. and Wermuth, N. *Multivariate Dependencies - Models, Analysis and Interpretation*. Chapman & Hall, 1996.
- 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.
- Kauermann, G. On a Dualization of Graphical Gaussian Models. *Scandinavian Journal of Statistics*, 23:106-116, 1996.
- Koster, J. T. A. Marginalizing and Conditioning in Graphical Models. *Bernoulli*, 8:817-840, 2002.
- 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.
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- Peña, J. M., Nilsson, R., Björkegren, J. and Tegnér, J. An Algorithm for Reading Dependencies from the Minimal Undirected Independence Map of a Graphoid that Satisfies Weak Transitivity. *Journal of Machine Learning Research*, 10:1071-1094, 2009.
- Peña, J. M. Towards Optimal Learning of Chain Graphs. arXiv:1109.5404 [stat.ML], 2011.
- Peña, J. M. Learning AMP Chain Graphs under Faithfulness. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, 251-258, 2012.
- Peña, J. M. Reading Dependencies from Covariance Graphs. *International Journal of Approximate Reasoning*, 54:216-227, 2013a.
- Peña, J. M. Learning AMP Chain Graphs and some Marginal Models Thereof under Faithfulness: Extended Version. arXiv:1303.0691[stat.ML], 2013b.
- Peña, J. M., Sonntag, D. and Nielsen, J. D. An Inclusion Optimal Algorithm for Chain Graph Structure Learning. Submitted, 2012.

- Richardson, T. and Spirtes, P. Ancestral Graph Markov Models. *The Annals of Statistics*, 30:962-1030, 2002.
- Roverato, A. and Studený, M. A Graphical Representation of Equivalence Classes of AMP Chain Graphs. *Journal of Machine Learning Research*, 7:1045-1078, 2006.
- 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.
- Sonntag, D. and Peña, J. M. Chain Graph Interpretations and their Relations. Submitted, 2013. Available at [www.ida.liu.se/~jospe/ecsqaru13extended.pdf](http://www.ida.liu.se/~jospe/ecsqaru13extended.pdf).
- 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, 1997a.
- Studený, M. On Marginalization, Collapsibility and Precollapsibility. In *Distributions with Given Marginals and Moment Problems*, 191-198. Kluwer, 1997b.
- Studený, M. *Probabilistic Conditional Independence Structures*. Springer, 2005.