Error AMP Chain Graphs

Jose M. Peña

Linköping University Post Print

Tweet

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

Error AMP Chain Graphs

Jose M. PEÑA

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

Abstract. Any regular Gaussian probability distribution that can be represented by an AMP chain graph (CG) can be expressed as a system of linear equations with correlated errors whose structure depends on the CG. However, the CG represents the errors implicitly, as no nodes in the CG correspond to the errors. We propose in this paper to add some deterministic nodes to the CG in order to represent the errors explicitly. We call the result an EAMP CG. We will show that, as desired, every AMP CG is Markov equivalent to its corresponding EAMP CG under marginalization of the error nodes. We will also show that every EAMP CG under marginalization of the error nodes is Markov equivalent to some LWF CG under marginalization of the error nodes, and that the latter is Markov equivalent to some directed and acyclic graph (DAG) under marginalization of the error nodes and conditioning on some selection nodes. This is important because it implies that the independence model represented by an AMP CG can be accounted for by some data generating process that is partially observed and has selection bias. Finally, we will show that EAMP CGs are closed under marginalization. This is a desirable feature because it guarantees parsimonious models under marginalization.

Keywords. probabilistic graphical models, directed and acyclic graphs, chain graphs

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. CGs extend Markov networks, i.e. undirected graphs, and Bayesian networks, i.e. directed and acyclic graphs (DAGs). Therefore, they can model symmetric and asymmetric relationships between the random variables of interest, which is one of the reasons of their popularity. However, unlike Markov and Bayesian networks whose interpretation is unique, there are four different interpretations of CGs as independence models [3,4,5,16]. In this paper, we are interested in the AMP interpretation [1,11] and the LWF interpretation [6,10].

Any regular Gaussian probability distribution that can be represented by an AMP CG can be expressed as a system of linear equations with correlated errors whose structure depends on the CG [1, Section 5]. However, the CG represents the errors implicitly, as no nodes in the CG correspond to the errors. We propose in this paper to add some deterministic nodes to the CG in order to represent the errors explicitly. We call the result an EAMP CG. We will show that, as desired, every AMP CG is Markov equivalent to its corresponding EAMP CG under marginalization of the error nodes, i.e. the independence model represented by the former coincides with the independence model represented by

the latter. We will also show that every EAMP CG under marginalization of the error nodes is Markov equivalent to some LWF CG under marginalization of the error nodes, and that the latter is Markov equivalent to some DAG under marginalization of the error nodes and conditioning on some selection nodes. The relevance of this result can be best explained by extending to AMP CGs what Koster stated for summary graphs [8, p. 838] and Richardson and Spirtes stated for ancestral graphs [13, p. 981]: The fact that an AMP CG has a DAG as departure point implies that the independence model associated with the former can be accounted for by some data generating process that is partially observed (corresponding to marginalization) and has selection bias (corresponding to conditioning). Finally, we will show that EAMP CGs are closed under marginalization, in the sense that every EAMP CG under marginalization of any superset of the error nodes is Markov equivalent to some EAMP CG under marginalization of the error nodes.[1] The relevance of this result can be best appreciated by noting that AMP CGs are not closed under marginalization [13, Section 9.4]. Therefore, the independence model represented by an AMP CG under marginalization may not be representable by any AMP CG. Therefore, we may have to represent it by an AMP CG with extra edges so as to avoid representing false independencies. However, if we consider the EAMP CG corresponding to the original AMP CG, then we will show that the marginal independence model can be represented by some EAMP CG under marginalization of the error nodes. The latter case is of course preferred, because the graphical model is more parsimonious as it does not include extra edges. See also [13, p. 965] for a discussion on the importance of the class of models considered being closed under marginalization.

It is worth mentioning that Andersson et al. have identified the conditions under which an AMP CG is Markov equivalent to some LWF CG [1, Theorem 6].[2] It is clear from these conditions that there are AMP CGs that are not Markov equivalent to any LWF CG. The results in this paper differ from those by Andersson et al., because we show that every AMP CG is Markov equivalent to some LWF CG with error nodes under marginalization of the error nodes.

It is also worth mentioning that Richardson and Spirtes show that there are AMP CGs that are not Markov equivalent to any DAG under marginalization and conditioning [13, p. 1025]. However, the results in this paper show that every AMP CG is Markov equivalent to some DAG with error and selection nodes under marginalization of the error nodes and conditioning of the selection nodes. Therefore, the independence model represented by any AMP CG has indeed some DAG as departure point and, thus, it can be accounted for by some data generating process. The results in this paper do not contradict those by Richardson and Spirtes, because they did not consider deterministic nodes while we do (recall that the error nodes are deterministic).

Finally, it is also worth mentioning that EAMP CGs are not the first graphical models to have DAGs as departure point or to be closed under marginalization. Specifically, summary graphs [4], MC graphs [8], ancestral graphs [13], and ribonless graphs [15]

---

[1]Our definition of closed under marginalization is an adaptation of the standard one to the fact that we only care about independence models under marginalization of the error nodes.

[2]To be exact, Andersson et al. have identified the conditions under which all and only the probability distributions that can be represented by an AMP CG can also be represented by some LWF CG. However, for any AMP or LWF CG G, there are Gaussian probability distributions that have all and only the independencies in the independence model represented by G, as shown in [11, Theorem 6.1] and [12, Theorems 1 and 2]. Then, our formulation is equivalent to the original formulation of the result by Andersson et al.

predate EAMP CGs and have the mentioned properties. However, none of these other classes of graphical models subsumes AMP CGs, i.e. there are independence models that can be represented by an AMP CG but not by any member of the other class [14, Section 4]. Therefore, none of these other classes of graphical models subsumes EAMP CGs under marginalization of the error nodes. This justifies the present study.

The rest of the paper is organized as follows. We start by reviewing some concepts in Section 1. We discuss in Section 2 the semantics of deterministic nodes in the context of AMP and LWF CGs. In Section 3, we introduce EAMP CGs and use them to show that every AMP CG is Markov equivalent to some LWF CG under marginalization. In that section we also show that every AMP CG is Markov equivalent to some DAG under marginalization and conditioning. In Section 4, we show that EAMP CGs are closed under marginalization. Finally, we close with some conclusions in Section 5.

1. Preliminaries

In this section, we review some concepts from 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 unless otherwise stated. The elements of V are not distinguished from singletons. The operators set union and set difference are given equal precedence in the expressions. The term maximal is always wrt set inclusion. All the graphs in this paper are simple, i.e. they contain at most one edge between any pair of nodes. Moreover, the edge is undirected or directed.

If a graph G contains an undirected or directed edge between two nodes V_1 and V_2, then we say that $V_1 - V_2$ or $V_1 \rightarrow V_2$ is in G. The parents of a set of nodes X of G is the set $pa_G(X) = \{V_1 | V_1 \rightarrow V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. A route between a node V_1 and a node V_n in G is a sequence of (not necessarily distinct) nodes $V_1, \ldots, V_n$ st $V_i - V_{i+1}$, $V_i \rightarrow V_{i+1}$ or $V_i \leftarrow V_{i+1}$ is in G for all $1 \leq i < n$. If the nodes in the route are all distinct, then the route is called a path. A route is called undirected if $V_i - V_{i+1}$ 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 strict ascendants of X is the set $san_G(X) = \{V_1 |$ there is a strictly descending route from V_1 to V_n in G, $V_1 \notin X$ and $V_n \in X\}$. A route $V_1, \ldots, V_n$ 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 $V_i \rightarrow V_{i+1}$ or $V_i - V_{i+1}$ is in G for all $1 < i < n$. A chain graph (CG) is a graph with no semidirected cycles. A set of nodes of a graph is connected if there exists an undirected path in the graph between every pair of nodes in the set. A connectivity component of a CG is a maximal connected set.

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

- every triplex node in ρ is in $Z \cup san_G(Z)$, and
- no non-triplex node B in ρ is in Z, unless $A - B - C$ is a subpath of ρ and some node in $pa_G(B)$ is not in Z.

A section of a route ρ in a CG is a maximal undirected subroute of ρ. A section $V_2 - \ldots - V_{n-1}$ of ρ is a collider section of ρ if $V_1 \rightarrow V_2 - \ldots - V_{n-1} \leftarrow V_n$ is a subroute of ρ. A route ρ in a CG is said to be Z-open when

- every collider section of ρ has a node in Z, and

- no non-collider section of ρ has a node in Z.

Let X, Y and Z denote three disjoint subsets of V. When there is no Z-open path (respectively route) in an AMP (respectively LWF) 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 $I_{AMP}(G)$ or $I_{LWF}(G)$, is the set of separations $X \perp_G Y | Z$. In general, $I_{AMP}(G) \neq I_{LWF}(G)$. However, if G is a directed and acyclic graph (DAG), then $I_{AMP}(G) = I_{LWF}(G)$. Given an AMP or LWF CG G and two disjoint subsets L and S of V, we denote by $[I(G)]_L^S$ the independence model represented by G under marginalization of the nodes in L and conditioning on the nodes in S. Specifically, $X \perp_G Y | Z$ is in $[I(G)]_L^S$ iff $X \perp_G Y | Z \cup S$ is in $I(G)$ and $X, Y, Z \subseteq V \smallsetminus L \smallsetminus S$.

Finally, we denote by $X \perp_p Y | Z$ that X is independent of Y given Z in a probability distribution p. We say that p is Markovian wrt an AMP or LWF CG G when $X \perp_p Y | Z$ if $X \perp_G Y | Z$ for all X, Y and Z disjoint subsets of V. We say that p is faithful to G when $X \perp_p Y | Z$ iff $X \perp_G Y | Z$ for all X, Y and Z disjoint subsets of V.

2. AMP and LWF CGs with Deterministic Nodes

We say that a node A of an AMP or LWF CG is determined by some $Z \subseteq V$ when $A \in Z$ or A is a function of Z. In that case, we also say that A is a deterministic node. We use $D(Z)$ to denote all the nodes that are determined by Z. From the point of view of the separations in an AMP or LWF CG, that a node is determined by but is not in the conditioning set of a separation has the same effect as if the node were actually in the conditioning set. We extend the definitions of separation for AMP and LWF CGs to the case where deterministic nodes may exist.

Given an AMP CG G, a path ρ in G is said to be Z-open when

- every triplex node in ρ is in $D(Z) \cup san_G(D(Z))$, and
- no non-triplex node B in ρ is in $D(Z)$, unless $A - B - C$ is a subpath of ρ and some node in $pa_G(B)$ is not in $D(Z)$.

Given an LWF CG G, a route ρ in G is said to be Z-open when

- every collider section of ρ has a node in $D(Z)$, and
- no non-collider section of ρ has a node in $D(Z)$.

It should be noted that we are not the first to consider graphical models with deterministic nodes. For instance, Geiger et al. consider DAGs with deterministic nodes. However, our definition of deterministic node is more general than theirs [7, Section 4].

3. From AMP CGs to DAGs Via EAMP CGs

Andersson et al. show that any regular Gaussian probability distribution p that is Markovian wrt an AMP CG G can be expressed as a system of linear equations with correlated errors whose structure depends on G [1, Section 5]. Specifically, assume without loss of generality that p has mean 0. Let K_i denote any connectivity component of G. Let $\Omega^i_{K_i, K_i}$ and $\Omega^i_{K_i, pa_G(K_i)}$ denote submatrices of the precision matrix Ω^i of $p(K_i, pa_G(K_i))$. Then, as shown in [2, Section 2.3.1], $K_i | pa_G(K_i) \sim \mathcal{N}(\beta^i pa_G(K_i), \Lambda^i)$

where $\beta^i = -(\Omega^i_{K_i,K_i})^{-1}\Omega^i_{K_i,pa_G(K_i)}$ and $(\Lambda^i)^{-1} = \Omega^i_{K_i,K_i}$. Then, p can be expressed as a system of linear equations with normally distributed errors whose structure depends on G as follows: $K_i = \beta^i\, pa_G(K_i) + \varepsilon^i$ where $\varepsilon^i \sim \mathcal{N}(0,\Lambda^i)$.

Note that for all $A,B \in K_i$ st $A-B$ is not in G, $A \perp_G B | pa_G(K_i) \cup K_i \smallsetminus A \smallsetminus B$ and thus $(\Lambda^i)^{-1}_{A,B} = 0$ [9, Proposition 5.2]. Note also that for all $A \in K_i$ and $B \in pa_G(K_i)$ st $A \leftarrow B$ is not in G, $A \perp_G B | pa_G(A)$ and thus $(\beta^i)_{A,B} = 0$. Let β_A contain the nonzero elements of the vector $(\beta^i)_{A,\bullet}$. Then, p can be expressed as a system of linear equations with correlated errors whose structure depends on G as follows: $A = \beta_A\, pa_G(A) + \varepsilon^A$ for all $A \in K_i$ where $covariance(\varepsilon^A,\varepsilon^B) = \Lambda^i_{A,B}$ for all $A,B \in K_i$.

It is worth mentioning that the mapping above between probability distributions and systems of linear equations is bijective [1, Section 5]. Note that no nodes in G correspond to the errors ε^A. Therefore, G represent the errors implicitly. We propose to represent them explicitly. This can easily be done by transforming G into what we call an EAMP CG G' as follows:

1   Let $G' = G$
2   For each node A in G
3   Add the node ε^A to G'
4   Add the edge $\varepsilon^A \to A$ to G'
5   For each edge $A-B$ in G
6   Add the edge $\varepsilon^A - \varepsilon^B$ to G'
7   Remove the edge $A-B$ from G'

The transformation above basically consists in adding the error nodes ε^A to G and connect them appropriately. Figure 1 shows an example. Note that every node $A \in V$ is determined by $pa_{G'}(A)$ and, what is more important in this paper, that ε^A is determined by $pa_{G'}(A) \smallsetminus \varepsilon^A \cup A$. Note also that, given $Z \subseteq V$, a node $A \in V$ is determined by Z iff $A \in Z$. The if part is trivial. To see the only if part, note that $\varepsilon^A \notin Z$ and thus A cannot be determined by Z unless $A \in Z$. Therefore, a node ε^A in G' is determined by Z iff $pa_{G'}(A) \smallsetminus \varepsilon^A \cup A \subseteq Z$ because, as shown, there is no other way for Z to determine $pa_{G'}(A) \smallsetminus \varepsilon^A \cup A$ which, in turn, determine ε^A. Let ε denote all the error nodes in G'. We define separation in EAMP CGs to be the same as separation in AMP CGs. In other words, an EAMP CG is interpreted as if it were an AMP CG. The following theorem confirms that the semantics of G' are as desired.

Theorem 1 $I_{AMP}(G) = [I_{AMP}(G')]^{\varnothing}_{\varepsilon}$.

Proof *It suffices to show that every Z-open path between α and β in G can be transformed into a Z-open path between α and β in G' and vice versa, with $\alpha, \beta \in V$ and $Z \subseteq V \smallsetminus \alpha \smallsetminus \beta$.*

Let ρ denote a Z-open path between α and β in G. We can easily transform ρ into a path ρ' between α and β in G': Simply, replace every maximal subpath of ρ of the form $V_1 - V_2 - \ldots - V_{n-1} - V_n$ ($n \geq 2$) with $V_1 \leftarrow \varepsilon^{V_1} - \varepsilon^{V_2} - \ldots - \varepsilon^{V_{n-1}} - \varepsilon^{V_n} \to V_n$. We now show that ρ' is Z-open.

First, if $B \in V$ is a triplex node in ρ', then ρ' must have one of the following subpaths:

$$A \longrightarrow B \longleftarrow C \quad A \longrightarrow B \longleftarrow \varepsilon^B - \varepsilon^C \quad \varepsilon^A - \varepsilon^B \to B \longleftarrow C$$

G G' G'' $[G']_{\{A,B,F\}}$

$$
\begin{array}{cccc}
A \longrightarrow B & \varepsilon^A \longrightarrow A \longrightarrow B \longleftarrow \varepsilon^B & \varepsilon^A \longrightarrow A \longrightarrow B \longleftarrow \varepsilon^B & \varepsilon^A \qquad\qquad \varepsilon^B \\
C \longrightarrow D & \varepsilon^C \longrightarrow C \qquad D \longleftarrow \varepsilon^D & \varepsilon^C \longrightarrow C \qquad D \longleftarrow \varepsilon^D & \varepsilon^C \longrightarrow C \qquad D \longleftarrow \varepsilon^D \\
 & & S_{\varepsilon^C \varepsilon^D} & \\
 & & S_{\varepsilon^C \varepsilon^E} \qquad S_{\varepsilon^D \varepsilon^F} & \\
 & & S_{\varepsilon^E \varepsilon^F} & \\
E \longrightarrow F & \varepsilon^E \longrightarrow E \qquad F \longleftarrow \varepsilon^F & \varepsilon^E \longrightarrow E \qquad F \longleftarrow \varepsilon^F & \varepsilon^E \longrightarrow E \qquad\qquad \varepsilon^F
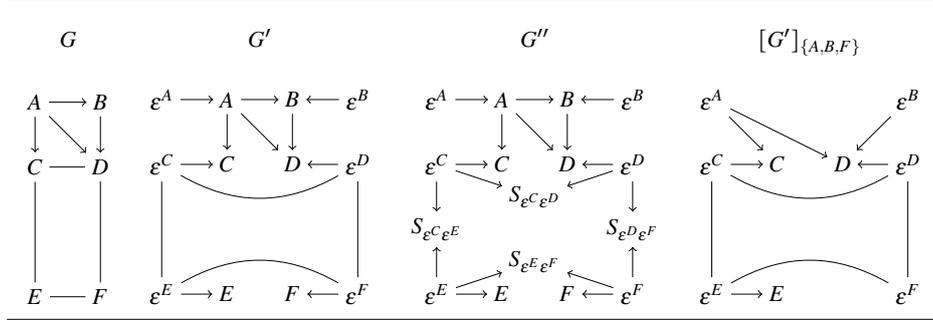\end{array}
$$

Figure 1. Example of the different transformations.

with $A, C \in V$. Therefore, ρ must have one of the following subpaths (specifically, if ρ' has the i-th subpath above, then ρ has the i-th subpath below):

$$A \longrightarrow B \longleftarrow C \quad A \longrightarrow B \longrightarrow C \quad A \longrightarrow B \longleftarrow C$$

In either case, B is a triplex node in ρ and, thus, $B \in Z \cup san_G(Z)$ for ρ to be Z-open. Then, $B \in Z \cup san_{G'}(Z)$ by construction of G' and, thus, $B \in D(Z) \cup san_{G'}(D(Z))$.

Second, if $B \in V$ is a non-triplex node in ρ', then ρ' must have one of the following subpaths:

$$A \longrightarrow B \longrightarrow C \quad A \longleftarrow B \longrightarrow C \quad A \longleftarrow B \longleftarrow C \quad A \longleftarrow B \longleftarrow \varepsilon^B \longrightarrow \varepsilon^C \quad \varepsilon^A \longrightarrow \varepsilon^B \longrightarrow B \longrightarrow C$$

with $A, C \in V$. Therefore, ρ must have one of the following subpaths (specifically, if ρ' has the i-th subpath above, then ρ has the i-th subpath below):

$$A \longrightarrow B \longrightarrow C \quad A \longleftarrow B \longrightarrow C \quad A \longleftarrow B \longleftarrow C \quad A \longleftarrow B \longrightarrow C \quad A \longrightarrow B \longrightarrow C$$

In either case, B is a non-triplex node in ρ and, thus, $B \notin Z$ for ρ to be Z-open. Since Z contains no error node, Z cannot determine any node in V that is not already in Z. Then, $B \notin D(Z)$.

Third, if ε^B is a non-triplex node in ρ' (note that ε^B cannot be a triplex node in ρ'), then ρ' must have one of the following subpaths:

$$A \longrightarrow B \longleftarrow \varepsilon^B \longrightarrow \varepsilon^C \quad \varepsilon^A \longrightarrow \varepsilon^B \longrightarrow B \longleftarrow C \quad \alpha = B \longleftarrow \varepsilon^B \longrightarrow \varepsilon^C \quad \varepsilon^A \longrightarrow \varepsilon^B \longrightarrow B = \beta$$

$$A \longleftarrow B \longleftarrow \varepsilon^B \longrightarrow \varepsilon^C \quad \varepsilon^A \longrightarrow \varepsilon^B \longrightarrow B \longrightarrow C \quad \varepsilon^A \longrightarrow \varepsilon^B \longrightarrow \varepsilon^C$$

with $A, C \in V$. Recall that $\varepsilon^B \notin Z$ because $Z \subseteq V \setminus \alpha \setminus \beta$. In the first case, if $\alpha = A$ then $A \notin Z$, else $A \notin Z$ for ρ to be Z-open. Then, $\varepsilon^B \notin D(Z)$. In the second case, if $\beta = C$ then $C \notin Z$, else $C \notin Z$ for ρ to be Z-open. Then, $\varepsilon^B \notin D(Z)$. In the third and fourth cases, $B \notin Z$ because $\alpha = B$ or $\beta = B$. Then, $\varepsilon^B \notin D(Z)$. In the fifth and sixth cases, $B \notin Z$ for ρ to be Z-open. Then, $\varepsilon^B \notin D(Z)$. The last case implies that ρ has the following subpath:

$$A \longrightarrow B \longrightarrow C$$

Thus, B is a non-triplex node in ρ, which implies that $B \notin Z$ or $pa_G(B) \smallsetminus Z \neq \varnothing$ for ρ to be Z-open. In either case, $\varepsilon^B \notin D(Z)$ (recall that $pa_{G'}(B) = pa_G(B) \cup \varepsilon^B$ by construction of G').

Finally, let ρ' denote a Z-open path between α and β in G'. We can easily transform ρ' into a path ρ between α and β in G: Simply, replace every maximal subpath of ρ' of the form $V_1 \leftarrow \varepsilon^{V_1} - \varepsilon^{V_2} - \ldots - \varepsilon^{V_{n-1}} - \varepsilon^{V_n} \rightarrow V_n$ $(n \geq 2)$ with $V_1 - V_2 - \ldots - V_{n-1} - V_n$. We now show that ρ is Z-open.

First, note that all the nodes in ρ are in V. Moreover, if B is a triplex node in ρ, then ρ must have one of the following subpaths:

$$A \longrightarrow B \longleftarrow C \quad A \longrightarrow B \longrightarrow C \quad A \longrightarrow B \longleftarrow C$$

with $A, C \in V$. Therefore, ρ' must have one of the following subpaths (specifically, if ρ has the i-th subpath above, then ρ' has the i-th subpath below):

$$A \longrightarrow B \longleftarrow C \quad A \longrightarrow B \longleftarrow \varepsilon^B - \varepsilon^C \quad \varepsilon^A - \varepsilon^B \longrightarrow B \longleftarrow C$$

In either case, B is a triplex node in ρ' and, thus, $B \in D(Z) \cup san_{G'}(D(Z))$ for ρ' to be Z-open. Since Z contains no error node, Z cannot determine any node in V that is not already in Z. Then, $B \in D(Z)$ iff $B \in Z$. Since there is no strictly descending route from B to any error node, then any strictly descending route from B to a node $D \in D(Z)$ implies that $D \in V$ which, as seen, implies that $D \in Z$. Then, $B \in san_{G'}(D(Z))$ iff $B \in san_{G'}(Z)$. Moreover, $B \in san_{G'}(Z)$ iff $B \in san_G(Z)$ by construction of G'. These results together imply that $B \in Z \cup san_G(Z)$.

Second, if B is a non-triplex node in ρ, then ρ must have one of the following subpaths:

$$A \longrightarrow B \longrightarrow C \quad A \longleftarrow B \longrightarrow C \quad A \longleftarrow B \longleftarrow C \quad A \longleftarrow B \longrightarrow C \quad A \longrightarrow B \longrightarrow C \quad A \longrightarrow B \longrightarrow C$$

with $A, C \in V$. Therefore, ρ' must have one of the following subpaths (specifically, if ρ has the i-th subpath above, then ρ' has the i-th subpath below):

$$A \longrightarrow B \longrightarrow C \quad A \longleftarrow B \longrightarrow C \quad A \longleftarrow B \longleftarrow C \quad A \longleftarrow B \longleftarrow \varepsilon^B - \varepsilon^C \quad \varepsilon^A - \varepsilon^B \longrightarrow B \longrightarrow C$$

$$\varepsilon^A - \varepsilon^B - \varepsilon^C$$

In the first five cases, B is a non-triplex node in ρ' and, thus, $B \notin D(Z)$ for ρ' to be Z-open. Since Z contains no error node, Z cannot determine any node in V that is not already in Z. Then, $B \notin Z$. In the last case, ε^B is a non-triplex node in ρ' and, thus, $\varepsilon^B \notin D(Z)$ for ρ' to be Z-open. Then, $B \notin Z$ or $pa_{G'}(B) \smallsetminus \varepsilon^B \smallsetminus Z \neq \varnothing$. Then, $B \notin Z$ or $pa_G(B) \smallsetminus Z \neq \varnothing$ (recall that $pa_{G'}(B) = pa_G(B) \cup \varepsilon^B$ by construction of G').

Theorem 2 *Assume that G' has the same deterministic relationships no matter whether it is interpreted as an AMP or LWF CG. Then, $I_{AMP}(G') = I_{LWF}(G')$.*

Proof *Assume for a moment that G' has no deterministic node. Note that G' has no induced subgraph of the form $A \to B - C$ with $A, B, C \in V \cup \varepsilon$. Such an induced subgraph is called a flag in [1, pp. 40-41]. There, the term biflag is also introduced. Its definition is irrelevant here. What is relevant here is the observation that a CG cannot have a biflag unless it has some flag. Therefore, G' has no biflags. Consequently, every probability distribution that is Markovian wrt G' when interpreted as an AMP CG is also Markovian wrt G' when interpreted as a LWF CG and vice versa [1, Corollary 1]. Now, note that there are Gaussian probability distributions that are faithful to G' when interpreted as an AMP CG [11, Theorem 6.1] as well as when interpreted as a LWF CG [12, Theorems 1 and 2]. Therefore, $I_{AMP}(G') = I_{LWF}(G')$. We denote this independence model by $I_{NDN}(G')$.*

Now, forget the momentary assumption made above that G' has no deterministic node. Recall that we assumed that $D(Z)$ is the same under the AMP and the LWF interpretations of G' for all $Z \subseteq V \cup \varepsilon$. Recall also that, from the point of view of the separations in an AMP or LWF CG, that a node is determined by the conditioning set has the same effect as if the node were in the conditioning set. Then, $X \perp_{G'} Y | Z$ is in $I_{AMP}(G')$ iff $X \perp_{G'} Y | D(Z)$ is in $I_{NDN}(G')$ iff $X \perp_{G'} Y | Z$ is in $I_{LWF}(G')$. Then, $I_{AMP}(G') = I_{LWF}(G')$.

The first major result of this paper is the following corollary, which shows that every AMP CG is Markov equivalent to some LWF CG under marginalization. The corollary follows from Theorems 1 and 2.

Corollary 1 $I_{AMP}(G) = [I_{LWF}(G')]_\varepsilon^\varnothing$.

Now, let G'' denote the DAG obtained from G' by replacing every edge $\varepsilon^A - \varepsilon^B$ in G' with $\varepsilon^A \to S_{\varepsilon^A \varepsilon^B} \leftarrow \varepsilon^B$. Figure 1 shows an example. The nodes $S_{\varepsilon^A \varepsilon^B}$ are called selection nodes. Let S denote all the selection nodes in G''. The following theorem relates the semantics of G' and G''.

Theorem 3 *Assume that G' and G'' have the same deterministic relationships. Then, $I_{LWF}(G') = [I_{LWF}(G'')]_\varnothing^S$.*

Proof *Assume for a moment that G' has no deterministic node. Then, G'' has no deterministic node either. We show below that every Z-open route between α and β in G' can be transformed into a $(Z \cup S)$-open route between α and β in G'' and vice versa, with $\alpha, \beta \in V \cup \varepsilon$. This implies that $I_{LWF}(G') = [I_{LWF}(G'')]_\varnothing^S$. We denote this independence model by $I_{NDN}(G')$.*

First, let ρ' denote a Z-open route between α and β in G'. Then, we can easily transform ρ' into a $(Z \cup S)$-open route ρ'' between α and β in G'': Simply, replace every edge $\varepsilon^A - \varepsilon^B$ in ρ' with $\varepsilon^A \to S_{\varepsilon^A \varepsilon^B} \leftarrow \varepsilon^B$. To see that ρ'' is actually $(Z \cup S)$-open, note that every collider section in ρ' is due to a subroute of the form $A \to B \leftarrow C$ with $A, B \in V$ and $C \in V \cup \varepsilon$. Then, any node that is in a collider (respectively non-collider) section of ρ' is also in a collider (respectively non-collider) section of ρ''.

Second, let ρ'' denote a $(Z \cup S)$-open route between α and β in G''. Then, we can easily transform ρ'' into a Z-open route ρ' between α and β in G': First, replace every subroute $\varepsilon^A \to S_{\varepsilon^A \varepsilon^B} \leftarrow \varepsilon^A$ of ρ'' with ε^A and, then, replace every subroute $\varepsilon^A \to S_{\varepsilon^A \varepsilon^B} \leftarrow \varepsilon^B$ of ρ'' with $\varepsilon^A - \varepsilon^B$. To see that ρ' is actually Z-open, note that every undirected edge in ρ' is between two noise nodes and recall that no noise node has incoming directed edges

in G'. Then, again every collider section in ρ' is due to a subroute of the form $A \to B \leftarrow C$ with $A, B \in V$ and $C \in V \cup \varepsilon$. Then, again any node that is in a collider (respectively non-collider) section of ρ' is also in a collider (respectively non-collider) section of ρ''.

Now, forget the momentary assumption made above that G' has no deterministic node. Recall that we assumed that $D(Z)$ is the same no matter whether we are considering G' or G'' for all $Z \subseteq V \cup \varepsilon$. Recall also that, from the point of view of the separations in a LWF CG, that a node is determined by the conditioning set has the same effect as if the node were in the conditioning set. Then, $X \perp_{G''} Y | Z$ is in $[I_{LWF}(G'')]_\varnothing^S$ iff $X \perp_{G'} Y | D(Z)$ is in $I_{NDN}(G')$ iff $X \perp_{G'} Y | Z$ is in $I_{LWF}(G')$. Then, $I_{LWF}(G') = [I_{LWF}(G'')]_\varnothing^S$.

The second major result of this paper is the following corollary, which shows that every AMP CG is Markov equivalent to some DAG under marginalization and conditioning. The corollary follows from Corollary 1, Theorem 3 and the fact that G'' is a DAG and, thus, $I_{AMP}(G'') = I_{LWF}(G'')$.

Corollary 2 $I_{AMP}(G) = [I_{LWF}(G'')]_\varepsilon^S = [I_{AMP}(G'')]_\varepsilon^S$.

4. EAMP CGs Are Closed under Marginalization

In this section, we show that EAMP CGs are closed under marginalization, meaning that for any EAMP CG G' and $L \subseteq V$ there is an EAMP CG $[G']_L$ st $[I_{AMP}(G')]_{L \cup \varepsilon} = [I_{AMP}([G']_L)]_\varepsilon$. We actually show how to transform G' into $[G']_L$.

To gain some intuition into the problem and our solution to it, assume that L contains a single node B. Then, marginalizing out B from the system of linear equations associated with G implies the following: For every C st $B \in pa_G(C)$, modify the equation $C = \beta_C \, pa_G(C) + \varepsilon^C$ by replacing B with the right-hand side of its corresponding equation, i.e. $\beta_B \, pa_G(B) + \varepsilon^B$ and, then, remove the equation $B = \beta_B \, pa_G(B) + \varepsilon^B$ from the system. In graphical terms, this corresponds to C inheriting the parents of B in G' and, then, removing B from G'. The following pseudocode formalizes this idea for any $L \subseteq V$.

1	Let $[G']_L = G'$
2	Repeat until all the nodes in L have been considered
3	Let B denote any node in L that has not been considered before
4	For each pair of edges $A \to B$ and $B \to C$ in $[G']_L$ with $A, C \in V \cup \varepsilon$
5	Add the edge $A \to C$ to $[G']_L$
6	Remove B and all the edges it participates in from $[G']_L$

Note that the result of the pseudocode above is the same no matter the ordering in which the nodes in L are selected in line 3. The following theorem confirms that the semantics of $[G']_L$ are as desired. The proof of the theorem can be found in the extended version of this paper, which is available on our website.

Theorem 4 $[I_{AMP}(G')]_{L \cup \varepsilon} = [I_{AMP}([G']_L)]_\varepsilon$.

5. Conclusions

In this paper, we have introduced EAMP CGs to model explicitly the errors in the system of linear equations associated to an AMP CG. We have shown that, as desired, every AMP CG is Markov equivalent to its corresponding EAMP CG under marginalization. We have used this result to show that every AMP CG is Markov equivalent to some LWF CG under marginalization. This result links the two most popular interpretations of CGs. We have used the previous result to show that every AMP CG is also Markov equivalent to some DAG under marginalization and conditioning. This result implies that the independence model represented by an AMP CG can be accounted for by some data generating process that is partially observed and has selection bias. Finally, we have shown that EAMP CGs are closed under marginalization, which guarantees parsimonious models under marginalization. We are currently studying how to modify EAMP CGs so that they become closed under conditioning too.

Acknowledgments

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] Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] Cox, D. R. and Wermuth, N. Linear Dependencies Represented by Chain Graphs. *Statistical Science*, 8:204-218, 1993.

[4] Cox, D. R. and Wermuth, N. *Multivariate Dependencies - Models, Analysis and Interpretation*. Chapman & Hall, 1996.

[5] Drton, M. Discrete Chain Graph Models. *Bernoulli*, 15:736-753, 2009.

[6] Frydenberg, M. The Chain Graph Markov Property. *Scandinavian Journal of Statistics*, 17:333-353 1990.

[7] Geiger, D., Verma, T. and Pearl, J. Identifying Independence in Bayesian Networks. *Networks*, 20:507-534, 1990.

[8] Koster, J. T. A. Marginalizing and Conditioning in Graphical Models. *Bernoulli*, 8:817-840, 2002.

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

[10] Lauritzen, S. L. and Wermuth, N. Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative. *Annual of Statistics*, 17:31-57, 1989.

[11] 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.

[12] Peña, J. M. Faithfulness in Chain Graphs: The Gaussian Case. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 588-599, 2011.

[13] Richardson, T. and Spirtes, P. Ancestral Graph Markov Models. *The Annals of Statistics*, 30:962-1030, 2002.

[14] Sadeghi, K. and Lauritzen, S. L. Markov Properties for Mixed Graphs. arXiv:1109.5909v4 [stat.OT].

[15] Sadeghi, K. Stable Mixed Graphs. *Bernoulli*, to appear.

[16] 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.