Journal of Machine Learning Research 10 (2009) 41-70 Submitted 7/07; Revised 9/08; Published 1/09

Markov Properties for Linear Causal Models with Correlated Errors

Changsung Kang CSKANG@CS.IASTATE.EDU
Jin Tian JTIAN@CS.IASTATE.EDU
Department of Computer Science
Iowa State University
Ames, IA 50011, USA

Editor: Andre Elisseeff

Abstract

A linear causal model with correlated errors, represented by a DAG with bi-directed edges, can be tested by the set of conditional independence relations implied by the model. A global Markov property specifies, by the d-separation criterion, the set of all conditional independence relations holding in any model associated with a graph. A local Markov property specifies a much smaller set of conditional independence relations which will imply all other conditional independence relations which hold under the global Markov property. For DAGs with bi-directed edges associated with arbitrary probability distributions, a local Markov property is given in Richardson (2003) which may invoke an exponential number of conditional independencies. In this paper, we show that for a class of linear structural equation models with correlated errors, there is a local Markov property which will invoke only a linear number of conditional independence relations. For general linear models, we provide a local Markov property that often invokes far fewer conditional independencies than that in Richardson (2003). The results have applications in testing linear structural equation models with correlated errors.

Keywords: Markov properties, linear causal models, linear structural equation models, graphical models

1. Introduction

Linear causal models called structural equation models (SEMs) are widely used for causal reasoning in social sciences, economics, and artificial intelligence (Goldberger, 1972; Bollen, 1989; Spirtes et al., 2001; Pearl, 2000). One important problem in the applications of linear causal models is testing a hypothesized model against the given data. While the conventional method involves maximum likelihood estimation of the covariance matrix, an alternative approach has been proposed recently which involves testing for the conditional independence relationships implied by the model (Spirtes et al., 1998; Pearl, 1998; Pearl and Meshkat, 1999; Pearl, 2000; Shipley, 2000, 2003). The advantages of using this new test method instead of the traditional global fitting test have been discussed in Pearl (1998), Shipley (2000), McDonald (2002) and Shipley (2003). The method can be applied in small data samples and it can test "local" features of the model.

To apply this test method, one needs to be able to identify the conditional independence relationships implied by an SEM. This can be achieved by representing the SEM with a graph called a path diagram (Wright, 1934) and then reading independence relations from the path diagram. For a linear SEM without correlated errors, the corresponding path diagram is a directed acyclic graph (DAG). The set of all conditional independence relations holding in any model associated with a

DAG, often called a global Markov property for the DAG, can be read by the d-separation criterion (Pearl, 1988). However, it is not necessary to test for all the independencies implied by the model as a subset of those independencies may imply all others. A local Markov property specifies a much smaller set of conditional independence relations which will imply (using the laws of probability) all other conditional independence relations that hold under the global Markov property. A well-known local Markov property for DAGs is that each variable is conditionally independent of its non-descendants given its parents (Lauritzen et al., 1990; Lauritzen, 1996). Based on this local Markov property, Pearl and Meshkat (1999) and Shipley (2000) proposed testing methods for linear SEMs without correlated errors that involve at most one conditional independence test for each pair of variables.

On the other hand, the path diagrams for linear SEMs with correlated errors are DAGs with bi-directed edges ($\leftrightarrow$) where bi-directed edges are used to represent correlated errors. A DAG with bi-directed edges is called an *acyclic directed mixed graph (ADMG)* in Richardson (2003). The set of all conditional independence relations encoded in an ADMG can still be read by (a natural extension of) the d-separation criterion (called m-separation in Richardson, 2003) which provides the global Markov property for ADMGs (Spirtes et al., 1998; Koster, 1999; Richardson, 2003). A local Markov property for ADMGs is given in Richardson (2003), which, in the worst case, may invoke an exponential number of conditional independence relations, a sharp difference with the local Markov property for DAGs, where only one conditional independence relation is associated with each variable. Shipley (2003) suggested a method for testing linear SEMs with correlated errors but the method may or may not, depending on the actual models, be able to find a subset of conditional independence relations that imply all others.

In this paper, we seek to improve the local Markov property given in Richardson (2003) for linear SEMs with correlated errors. The local Markov property in Richardson (2003) is applicable for ADMGs associated with arbitrary probability distributions. Specifically, only semi-graphoid axioms which must hold in all probability distributions (Pearl, 1988) are used in showing that the set of conditional independence relations specified by the local Markov property will imply all those specified by the global Markov property. On the other hand, in linear SEMs, variables are assumed to have normal distributions, and it is known that normal distributions also satisfy the so-called composition axiom. Therefore, in this paper, we look for local Markov properties for ADMGs associated with probability distributions that satisfy the composition axiom. We will show that for a class of ADMGs, the local Markov property will invoke only one conditional independence relation for each variable, and therefore testing for the corresponding linear SEMs will involve at most one conditional independence test for each pair of variables. For general ADMGs, we provide a procedure that reduces the number of conditional independencies invoked by the local Markov property given in Richardson (2003), and therefore reduces the complexity of testing linear SEMs with correlated errors.

In the test of conditional independence relations, the efficiency of the test is influenced by the size of the conditioning set (that is, the number of conditioning variables) with a small conditioning set having advantage over a large one. The conditional independence relations invoked by the standard local Markov property for DAGs use a parent set as the conditioning set. Pearl and Meshkat (1999) have shown for linear SEMs without correlated errors how to find a set of conditional independence relations that may involve fewer conditioning variables. In this paper, we also generalize this result to linear SEMs with correlated errors.

The paper is organized as follows. In Section 2, we introduce linear SEMs, give basic notation and definitions, and present the local Markov property developed in Richardson (2003). In Section 3, we show that for a class of ADMGs, there is a local Markov property for probability distributions satisfying the composition axiom that invokes only a linear number of conditional independence relations. We also show a local Markov property that may involve fewer conditioning variables. In Section 4, we consider general ADMGs (for probability distributions satisfying the composition axiom) and show a local Markov property that invokes fewer conditional independencies than that in Richardson (2003). Section 5 concludes the paper.

2. Preliminaries and Motivation

In this section, we give basic definitions and introduce some relevant concepts.

2.1 Linear Causal Models

The SEM technique was developed by geneticists (Wright, 1934) and economists (Haavelmo, 1943) for assessing cause-effect relationships from a combination of statistical data and qualitative causal assumptions. It is an important causal analysis tool widely used in social sciences, economics, and artificial intelligence (Goldberger, 1972; Duncan, 1975; Bollen, 1989; Spirtes et al., 2001). For a review of SEMs and causality we refer to Pearl (1998).

In an SEM, the causal relationships among a set of variables are often assumed to be linear and expressed by linear equations. Each equation describes the dependence of one variable in terms of the others. For example, an equation

$$Y = aX + \varepsilon \tag{1}$$

represents that X may have a *direct* causal influence on Y and that no other variables have (direct) causal influences on Y except those factors (represented by the error term ε traditionally assumed to have normal distribution) that are omitted from the model. The parameter a quantifies the (direct) causal effect of X on Y. An equation like (1) with a causal interpretation represents an autonomous causal mechanism and is said to be *structural*.

As an example, consider the following model from Pearl (2000) that concerns the relations between smoking (X) and lung cancer (Y), mediated by the amount of tar (Z) deposited in a person's lungs:

$$\begin{aligned} X &= \varepsilon_1, \\ Z &= aX + \varepsilon_2, \\ Y &= bZ + \varepsilon_3. \end{aligned}$$

The model assumes that the amount of tar deposited in the lungs depends on the level of smoking (and external factors) and that the production of lung cancer depends on the amount of tar in the lungs but smoking has no effect on lung cancer except as mediated through tar deposits. To fully specify the model, we also need to decide whether those omitted factors (ε_1, ε_2, ε_3) are correlated or not. We may assume that no other factor that affects tar deposit is correlated with the omitted factors that affect smoking or lung cancer ($Cov(\varepsilon_1, \varepsilon_2) = Cov(\varepsilon_2, \varepsilon_3) = 0$). However, there might be unobserved factors (say some unknown carcinogenic genotype) that affect both smoking and lung

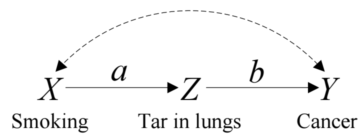


Figure 1: Causal diagram illustrating the effect of smoking on lung cancer

cancer ($Cov(\varepsilon_1, \varepsilon_3) \neq 0$), but the genotype nevertheless has no effect on the amount of tar in the lungs except indirectly (through smoking). Often, it is illustrative to express our qualitative causal assumptions in terms of a graphical representation, as shown in Figure 1.

We now formally define the model that we will consider in this paper. A *linear causal model* (or *linear SEM*) over a set of random variables $V = \{V_1, \ldots, V_n\}$ is given by a set of structural equations of the form

$$V_j = \sum_i c_{ji} V_i + \varepsilon_j, \; j = 1, \ldots, n, \tag{2}$$

where the summation is over the variables in V judged to be immediate causes of V_j. c_{ji}, called a *path coefficient*, quantifies the direct causal influence of V_i on V_j. ε_j's represent "error" terms due to omitted factors and are assumed to have normal distribution. We consider recursive models and assume that the summation in (2) is for $i < j$, that is, $c_{ji} = 0$ for $i \geq j$.

We denote the covariances between observed variables $\sigma_{ij} = Cov(V_i, V_j)$, and between error terms $\psi_{ij} = Cov(\varepsilon_i, \varepsilon_j)$. We denote the following matrices, $\Sigma = [\sigma_{ij}]$, $\Psi = [\psi_{ij}]$, and $C = [c_{ij}]$. The parameters of the model are the non-zero entries in the matrices C and Ψ. A parameterization of the model assigns a value to each parameter in the model, which then determines a unique covariance matrix Σ given by (see, for example, Bollen, 1989)

$$\Sigma = (I - C)^{-1} \Psi ((I - C)^t)^{-1}.$$

The structural assumptions encoded in the model are the zero path coefficients and zero error covariances. The model structure can be represented by a DAG G with (dashed) bi-directed edges (an ADMG), called a *causal diagram* (or *path diagram*), as follows: the nodes of G are the variables $V_1, \ldots, V_n$; there is a directed edge from V_i to V_j in G if V_i appears in the structural equation for V_j, that is, $c_{ji} \neq 0$; there is a bi-directed edge between V_i and V_j if the error terms ε_i and ε_j have non-zero correlation. For example, the smoking-and-lung-cancer SEM is represented by the causal diagram in Figure 1, in which each directed edge is annotated by the corresponding path coefficient.

We note that linear SEMs are often used without explicit causal interpretation. A linear SEM in which error terms are uncorrelated consists of a set of regression equations. Note that an equation as given by (2) is a regression equation if and only if ε_j is uncorrelated with each V_i ($Cov(V_i, \varepsilon_j) = 0$). Hence, an equation in an SEM with correlated errors may not be a regression equation. Linear SEMs provide a more powerful way to model data than the regression models taking into account correlated error terms.

2.2 Model Testing and Markov Properties

One important task in the applications of linear SEMs is to test a model against data. One approach for this task is to test for the conditional independence relationships implied by the model, which can be read from the causal diagram by the d-separation criterion as defined in the following.[1] A *path* between two vertices V_i and V_j in an ADMG consists of a sequence of consecutive edges of any type (directed or bi-directed). A vertex V_i is said to be an *ancestor* of a vertex V_j if there is a path $V_i \to \cdots \to V_j$. A non-endpoint vertex W on a path is called a *collider* if two arrowheads on the path meet at W, that is, $\to W \leftarrow$, $\leftrightarrow W \leftrightarrow$, $\leftrightarrow W \leftarrow$, $\to W \leftrightarrow$; all other non-endpoint vertices on a path are *non-colliders*, that is, $\leftarrow W \to$, $\leftarrow W \leftarrow$, $\to W \to$, $\leftrightarrow W \to$, $\leftarrow W \leftrightarrow$. A path between vertices V_i and V_j in an ADMG is said to be *d-connecting given a set* of vertices Z if

1. every non-collider on the path is not in Z, and

2. every collider on the path is an ancestor of a vertex in Z.

If there is no path d-connecting V_i and V_j given Z, then V_i and V_j are said to be *d-separated* given Z. Sets X and Y are said to be *d-separated* given Z, if for every pair V_i, V_j, with $V_i \in X$ and $V_j \in Y$, V_i and V_j are d-separated given Z. Let $I(X,Z,Y)$ denote that X is conditionally independent of Y given Z. The set of all the conditional independence relations encoded by a causal diagram G is specified by the following global Markov property.

Definition 1 (The Global Markov Property (GMP)) *A probability distribution P is said to satisfy the global Markov property for G if for arbitrary disjoint sets X,Y,Z with X and Y being nonempty,*

$$\text{(GMP)} \qquad X \textit{ is d-separated from } Y \textit{ given } Z \textit{ in } G \Longrightarrow I(X,Z,Y).$$

The global Markov property typically involves a vast number of conditional independence relations and it is possible to test for a subset of those independencies that will imply all others. A local Markov property specifies a much smaller set of conditional independence relations which will imply by the laws of probability all other conditional independence relations that hold under the global Markov property. For example, a well-known local Markov property for DAGs is that each variable is conditionally independent of its non-descendants given its parents. The causal diagram for a linear SEM with correlated errors is an ADMG and a local Markov property for ADMGs is given in Richardson (2003).

Note that in linear SEMs, the conditional independence relations will correspond to zero partial correlations (Lauritzen, 1996):

$$\rho_{V_iV_j.Z} = 0 \Longleftrightarrow I(\{V_i\},Z,\{V_j\}).$$

As an example, for the linear SEM with the causal diagram in Figure 2, if we use the local Markov property in Richardson (2003), then we need to test for the vanishing of the following set of partial correlations (for ease of notation, we write $\rho_{ij.Z}$ to denote $\rho_{V_iV_j.Z}$):

$$\{\rho_{21}, \rho_{32.1}, \rho_{43.2}, \rho_{41.2}, \rho_{54.3}, \rho_{52.3}, \rho_{51.3}, \rho_{64.53}, \rho_{62.53}, \rho_{61.53}, \rho_{64.3}, \rho_{62.3}, \rho_{61.3}, \rho_{72.6543}, \\ \rho_{71.6543}, \rho_{72.643}, \rho_{71.643}, \rho_{75.4}, \rho_{73.4}, \rho_{72.4}, \rho_{71.4}\}. \tag{3}$$

The local Markov property in Richardson (2003) is valid for any probability distributions. In fact, the equivalence of the global and local Markov properties is proved using the following so-called *semi-graphoid axioms* (Pearl, 1988) that probabilistic conditional independencies must satisfy:

1. The d-separation criterion was originally defined for DAGs (Pearl, 1988) but can be naturally extended for ADMGs and is called m-separation in Richardson (2003).

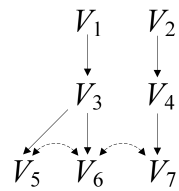


Figure 2: A causal diagram

- Symmetry
$$I(X,Z,Y) \Longleftrightarrow I(Y,Z,X).$$
- Decomposition
$$I(X,Z,Y \cup W) \Longrightarrow I(X,Z,Y) \ \& \ I(X,Z,W).$$
- Weak Union
$$I(X,Z,Y \cup W) \Longrightarrow I(X,Z \cup W,Y).$$
- Contraction
$$I(X,Z,Y) \ \& \ I(X,Z \cup Y,W) \Longrightarrow I(X,Z,Y \cup W).$$

where X, Y, Z, and W are disjoint sets of variables.

On the other hand, in linear SEMs the variables are assumed to have normal distributions, and normal distributions also satisfy the following *composition* axiom:

- Composition
$$I(X,Z,Y) \ \& \ I(X,Z,W) \Longrightarrow I(X,Z,Y \cup W).$$

Therefore, we expect a local Markov property for linear SEMs to invoke fewer conditional independence relations than that for arbitrary distributions. In this paper, we will derive reduced local Markov properties for linear SEMs by making use of the composition axiom. As an example, for the linear SEM in Figure 2, a local Markov property which we will present in this paper (see Section 3.3) says that we only need to test for the vanishing of the following set of partial correlations:

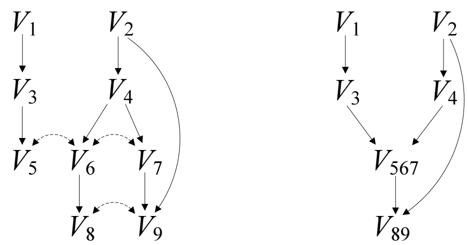
$$\{\rho_{21},\rho_{32},\rho_{43},\rho_{41},\rho_{54},\rho_{52},\rho_{51.3},\rho_{64},\rho_{62},\rho_{61.3},\rho_{75},\rho_{73},\rho_{71},\rho_{72.4}\}.$$

The number of tests needed and the size of the conditioning set Z are both substantially reduced compared with (3), thus leading to a more economical way of testing the given model.

2.3 A Local Markov Property for ADMGs

In this section, we describe the local Markov property for ADMGs associated with arbitrary probability distributions presented in Richardson (2003). In this paper, this Markov property will be used as an important tool to prove the equivalence of our local Markov properties and the global Markov property.

First, we define some graphical notations. For a vertex X in an ADMG G, $\text{pa}_G(X) \equiv \{Y|Y \to X$ in $G\}$ is the set of *parents* of X. $\text{sp}_G(X) \equiv \{Y|Y \leftrightarrow X$ in $G\}$ is the set of *spouses* of X. $\text{an}_G(X) \equiv$



(a) (b)

Figure 3: An ADMG and its compressed graph

$\{Y|Y \to \cdots \to X$ in G or $Y = X\}$ is the set of *ancestors* of X. And $\mathrm{de}_G(X) \equiv \{Y|Y \leftarrow \cdots \leftarrow X$ in G or $Y = X\}$ is the set of *descendants* of X. These definitions will be applied to sets of vertices, so that, for example, $\mathrm{pa}_G(A) \equiv \cup_{X\in A}\mathrm{pa}_G(X)$, $\mathrm{sp}_G(A) \equiv \cup_{X\in A}\mathrm{sp}_G(X)$, etc.

Definition 2 (C-component) *A c-component of G is a maximal set of vertices in G such that any two vertices in the set are connected by a path on which every edge is of the form $\leftrightarrow$; a vertex that is not connected to any bi-directed edge forms a c-component by itself.*

For example, the ADMG in Figure 3 (a) is composed of 6 c-components $\{V_1\}$, $\{V_2\}$, $\{V_3\}$, $\{V_4\}$, $\{V_5, V_6, V_7\}$ and $\{V_8, V_9\}$. The *district* of X in G is the c-component of G that includes X. Thus,

$$\mathrm{dis}_G(X) \equiv \{Y|Y \leftrightarrow \cdots \leftrightarrow X \text{ in } G \text{ or } Y = X\}.$$

For example, in Figure 3 (a), we have $\mathrm{dis}_G(V_5) = \{V_5, V_6, V_7\}$ and $\mathrm{dis}_G(V_8) = \{V_8, V_9\}$. A set A is said to be *ancestral* if it is closed under the ancestor relation, that is, if $\mathrm{an}_G(A) = A$. Let G_A denote the induced subgraph of G on the vertex set A, formed by removing from G all vertices that are not in A, and all edges that do not have both endpoints in A.

Definition 3 (Markov Blanket)[2] *If A is an ancestral set in an ADMG G, and X is a vertex in A that has no children in A then the* Markov blanket of vertex X with respect to the induced subgraph on A, *denoted* $\mathrm{mb}(X,A)$ *is defined to be*

$$\mathrm{mb}(X,A) \equiv \mathrm{pa}_{G_A}(\mathrm{dis}_{G_A}(X)) \cup (\mathrm{dis}_{G_A}(X) \setminus \{X\}).$$

For example, for an ancestral set $A = \mathrm{an}_G(\{V_5, V_6\}) = \{V_1, V_2, V_3, V_4, V_5, V_6\}$ in Figure 3 (a), we have

$$\mathrm{mb}(V_5, A) = \{V_3, V_4, V_6\}.$$

An ordering ($\prec$) on the vertices of G is said to be consistent with G if $X \prec Y \Rightarrow Y \notin \mathrm{an}_G(X)$. Given a consistent ordering $\prec$, let $\mathrm{pre}_{G,\prec}(X) \equiv \{Y|Y \prec X \text{ or } Y = X\}$.

Definition 4 (The Ordered Local Markov Property (LMP,$\prec$)) *A probability distribution P satisfies the ordered local Markov property for G with respect to a consistent ordering $\prec$, if, for any X and ancestral set A such that $X \in A \subseteq \mathrm{pre}_{G,\prec}(X)$,*

$$\text{(LMP,}\prec\text{)} \qquad I(\{X\}, \mathrm{mb}(X,A), A \setminus (\mathrm{mb}(X,A) \cup \{X\})). \tag{4}$$

2. The definition of Markov blanket here follows that in Richardson (2003) and is compatible with that in Pearl (1988).

Theorem 5 (Richardson, 2003) *If G is an ADMG and $\prec$ is a consistent ordering, then a probability distribution P satisfies the ordered local Markov property for G with respect to $\prec$ if and only if P satisfies the global Markov property for G.*

We will write (GMP) $\Longleftrightarrow$ (LMP,$\prec$) to denote the equivalence of the two Markov properties. Therefore the (smaller) set of conditional independencies specified in the ordered local Markov property will imply all other conditional independencies which hold under the global Markov property. It is possible to further reduce the number of conditional independence relations in the ordered local Markov property. An ancestral set A, with $X \in A \subseteq \mathrm{pre}_{G,\prec}(X)$ is said to be *maximal with respect to the Markov blanket* $\mathrm{mb}(X,A)$ if, whenever there is a set B such that $A \subseteq B \subseteq \mathrm{pre}_{G,\prec}(X)$ and $\mathrm{mb}(X,A) = \mathrm{mb}(X,B)$, then $A = B$. For example, suppose that we are given an ordering $\prec$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$ for the graph G in Figure 3 (a). While an ancestral set $A = \mathrm{an}_G(\{V_3,V_6,V_7\}) = \{V_1,V_2,V_3,V_4,V_6,V_7\}$ is maximal with respect to the Markov blanket $\mathrm{mb}(V_7,A) = \{V_4,V_6\}$, an ancestral set $A' = \mathrm{an}_G(\{V_6,V_7\}) = \{V_2,V_4,V_6,V_7\}$ is not. It was shown that we only need to consider ancestral sets A which are maximal with respect to $\mathrm{mb}(X,A)$ in the ordered local Markov property (Richardson, 2003). Thus, we will consider only maximal ancestral sets A when we discuss (LMP,$\prec$) for the rest of this paper. The following lemma characterizes maximal ancestral sets.

Lemma 6 (Richardson, 2003) *Let X be a vertex and A an ancestral set in G with consistent ordering $\prec$ such that $X \in A \subseteq \mathrm{pre}_{G,\prec}(X)$. The set A is maximal with respect to the Markov blanket* mb*(X,A) if and only if*

$$A = \mathrm{pre}_{G,\prec}(X) \setminus \mathrm{de}_G(\mathrm{h}(X,A))$$

where

$$\mathrm{h}(X,A) \equiv \mathrm{sp}_G\Big(\mathrm{dis}_{G_A}(X)\Big) \setminus \Big(\{X\} \cup \mathrm{mb}(X,A)\Big).$$

Even though we only consider maximal ancestral sets, the ordered local Markov property may still invoke an exponential number of conditional independence relations. For example, for a vertex X, if $\mathrm{dis}_G(X) \subseteq \mathrm{pre}_{G,\prec}(X)$ and $\mathrm{dis}_G(X)$ has a clique of n vertices joined by bi-directed edges, then there are at least $O(2^{n-1})$ different Markov blankets.

It should be noted that only the semi-graphoid axioms were used to prove Theorem 5 on the equivalence of the two Markov properties and no assumptions about probability distributions were made. Next we will show that the ordered local Markov property can be further reduced if we use the composition axiom in addition to the semi-graphoid axioms. The local Markov properties we obtained (in Sections 3 and 4) are not restricted to linear causal models in that they are actually valid for any probability distributions that satisfy the composition axiom.

3. Markov Properties for ADMGs without Directed Mixed Cycles

In this section, we introduce three local Markov properties for a class of ADMGs and show that they are equivalent to the global Markov property. Also, we discuss related work in maximal ancestral graphs and chain graphs. First, we give some definitions.

Definition 7 **(Directed Mixed Cycle)** *A path is said to be a directed mixed path from X to Y if it contains at least one directed edge and every edge on the path is either of the form $Z \leftrightarrow W$, or $Z \to W$ with W between Z and Y. A directed mixed path from X to Y together with an edge $Y \to X$ or $Y \leftrightarrow X$ is called a directed mixed cycle.*

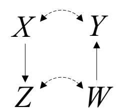


Figure 4: Directed mixed cycles

For example, the path $X \to Z \leftrightarrow W \to Y \leftrightarrow X$ in the graph in Figure 4 forms a directed mixed cycle. In this section, we will consider only ADMGs without directed mixed cycles.

Definition 8 (Compressed Graph) *Let G be an ADMG. The compressed graph of G is defined to be the graph $G' = (V', E')$, $V' = \{V_C \mid C$ is a c-component of $G\}$, $E' = \{V_{C_i} \to V_{C_j} \mid$ there is an edge $X \to Y$ in G such that $X \in C_i, Y \in C_j\}$.*

Figure 3 shows an ADMG and its compressed graph. If there exists a directed mixed cycle in an ADMG G, there will be a cycle or a self-loop in the compressed graph of G. For example, if for two vertices X and Y in a c-component C of G there exists an edge $X \to Y$, then the compressed graph of G contains a self-loop $\overset{\curvearrowright}{V_C}$. The following proposition holds.

Proposition 9 *Let G be an ADMG. The compressed graph of G is a DAG if and only if G has no directed mixed cycles.*

3.1 The Reduced Local Markov Property

In this section, we introduce a local Markov property for ADMGs without directed mixed cycles which only invokes a linear number of conditional independence relations and show that it is equivalent to the global local Markov property.

Definition 10 (The Reduced Local Markov Property (RLMP)) *Let G be an ADMG without directed mixed cycles. A probability distribution P is said to satisfy the reduced local Markov property for G if*

$$\text{(RLMP)} \qquad \forall X \in V, \quad I(\{X\}, \mathrm{pa}_G(X), V \setminus \mathrm{f}(X, G)) \tag{5}$$

where $\mathrm{f}(X,G) \equiv \mathrm{pa}_G(X) \cup \mathrm{de}_G(\{X\} \cup \mathrm{sp}_G(X))$.

The reduced local Markov property states that *a variable is independent of the variables that are neither its descendants nor its spouses' descendants given its parents.*

Theorem 11 *If a probability distribution P satisfies the composition axiom and an ADMG G has no directed mixed cycles, then*

$$\text{(GMP)} \Longleftrightarrow \text{(RLMP)}.$$

Proof: (GMP) $\Longrightarrow$ (RLMP)
We need to prove that any variable X is d-separated from $V \setminus \mathrm{f}(X,G)$ given $\mathrm{pa}_G(X)$ in G with no directed mixed cycle. Consider a vertex $\alpha \in V \setminus \mathrm{f}(X,G)$. We will show that there is no path d-connecting X and α given $\mathrm{pa}_G(X)$. There are four possible cases for any path between X and α.

1. $X \leftarrow \beta \cdots \alpha$
2. $X \rightarrow \cdots \rightarrow \delta \leftarrow\!\!* \cdots \alpha$
3. $X \leftrightarrow \gamma \leftarrow\!\!* \cdots \alpha$
4. $X \leftrightarrow \gamma \rightarrow \cdots \rightarrow \delta \leftarrow\!\!* \cdots \alpha$

A symbol $*$ serves as a wildcard for an end of an edge. For example, $\leftarrow\!\!*$ represents both $\leftarrow$ and $\leftrightarrow$. In case 1, $\beta \in \mathrm{pa}_G(X)$. In case 2, the collider δ is not an ancestor of a vertex in $\mathrm{pa}_G(X)$ (otherwise, there would be a cycle). In cases 3 and 4, neither γ nor δ is an ancestor of a vertex in $\mathrm{pa}_G(X)$ (otherwise, there would be directed mixed cycles). In any case, the path is not d-connecting given $\mathrm{pa}_G(X)$. ∎

Proof: $(\mathrm{RLMP}) \Longrightarrow (\mathrm{GMP})$
We will show that for some consistent ordering $\prec$, $(\mathrm{RLMP}) \Longrightarrow (\mathrm{LMP}, \prec)$. Then, by Theorem 5, we have $(\mathrm{RLMP}) \Longrightarrow (\mathrm{GMP})$.

We construct a consistent ordering with the desired property as follows.

1. Construct the compressed graph G' of G.
2. Let $\prec'$ be any consistent ordering on G'. Construct a consistent ordering $\prec$ from $\prec'$ by replacing each V_C (corresponding to each c-component C of G) in $\prec'$ with the vertices in C (the ordering of the vertices in C is arbitrary).

We now prove that (RLMP) $\Longrightarrow$ (LMP,$\prec$). Assume that a probability distribution P satisfies (RLMP). Consider the set of conditional independence relations invoked by (LMP,$\prec$) for each variable X given in (4). First, observe that for any vertex Y in $\mathrm{dis}_{G_A}(X)$, we have

$$A \setminus (\mathrm{pa}_G(Y) \cup \{Y\} \cup \mathrm{sp}_G(Y)) \subseteq V \setminus \mathrm{f}(Y,G),$$

since

$$\begin{aligned}
&A \setminus (\mathrm{pa}_G(Y) \cup \{Y\} \cup \mathrm{sp}_G(Y)) \\
&= A \setminus \Big(\big(\mathrm{pa}_G(Y) \cup \{Y\} \cup \mathrm{sp}_G(Y)\big) \cup \big(\mathrm{de}_G(\{Y\} \cup \mathrm{sp}_G(Y)) \setminus (\{Y\} \cup \mathrm{sp}_G(Y))\big)\Big) \qquad (6) \\
&= A \setminus \mathrm{f}(Y,G).
\end{aligned}$$

The equality (6) holds since the vertices in $\mathrm{de}_G(\{Y\} \cup \mathrm{sp}_G(Y)) \setminus (\{Y\} \cup \mathrm{sp}_G(Y))$ do not appear in A (because of the way $\prec$ is constructed, no descendant of $\mathrm{dis}_{G_A}(X)$ is in A). Thus, by (5), for all Y in $\mathrm{dis}_{G_A}(X)$, we have

$$I(\{Y\}, \mathrm{pa}_G(Y), A \setminus (\mathrm{pa}_G(Y) \cup \{Y\} \cup \mathrm{sp}_{G_A}(Y))).$$

Let $S_1 = \mathrm{pa}_G(\mathrm{dis}_{G_A}(X)) \setminus \mathrm{pa}_G(Y)$ and $S_2 = A \setminus (\mathrm{mb}(X,A) \cup \{X\})$. It follows that

$$\begin{aligned}
&S_1 \subseteq A \setminus (\mathrm{pa}_G(Y) \cup \{Y\} \cup \mathrm{sp}_G(Y)) \text{ and} \\
&S_2 \subseteq A \setminus (\mathrm{pa}_G(Y) \cup \{Y\} \cup \mathrm{sp}_G(Y)).
\end{aligned}$$

Also, we have

$$S_1 \cap S_2 = \emptyset,$$

since $S_1 \subseteq \text{mb}(X,A)$. Therefore, for $Y \in \text{dis}_{G_A}(X)$,

$$\begin{array}{ll}
I(\{Y\}, \text{pa}_G(Y), S_1 \cup S_2) & \text{by decomposition} \\
I(\{Y\}, \text{pa}_G(Y) \cup S_1, S_2) & \text{by weak union} \\
I(\text{dis}_{G_A}(X), \text{pa}_G(\text{dis}_{G_A}(X)), A \setminus (\text{mb}(X,A) \cup \{X\})) & \text{by composition} \\
I(\{X\}, \text{pa}_G(\text{dis}_{G_A}(X)) \cup (\text{dis}_{G_A}(X) \setminus \{X\}), & \\
\quad A \setminus (\text{mb}(X,A) \cup \{X\})) & \text{by weak union.}
\end{array}$$

Thus, we have

$$I(\{X\}, \text{mb}(X,A), A \setminus (\text{mb}(X,A) \cup \{X\}))$$

by the definition of the Markov blanket of X with respect to A. ■

As an example, consider the ADMG G in Figure 3 (a) which has no directed mixed cycles. The graph in Figure 3 (b) is the compressed graph G' of G described in the proof. From the ordering $\prec'$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_{567} \prec V_{89}$, we obtain the ordering $\prec$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$. The ordered local Markov property (LMP,$\prec$) involves the following conditional independence relations:

$$\begin{array}{ll}
I(\{V_2\}, \emptyset, \{V_1\}), & I(\{V_3\}, \{V_1\}, \{V_2\}), \\
I(\{V_4\}, \{V_2\}, \{V_1, V_3\}), & I(\{V_5\}, \{V_3\}, \{V_1, V_2, V_4\}), \\
I(\{V_6\}, \{V_3, V_4, V_5\}, \{V_1, V_2\}), & I(\{V_6\}, \{V_4\}, \{V_1, V_2, V_3\}), \\
I(\{V_7\}, \{V_3, V_4, V_5, V_6\}, \{V_1, V_2\}), & I(\{V_7\}, \{V_4, V_6\}, \{V_1, V_2, V_3\}), \\
I(\{V_7\}, \{V_4\}, \{V_1, V_2, V_3, V_5\}), & I(\{V_8\}, \{V_6\}, \{V_1, V_2, V_3, V_4, V_5, V_7\}), \\
I(\{V_9\}, \{V_2, V_6, V_7, V_8\}, \{V_1, V_3, V_4, V_5\}), & I(\{V_9\}, \{V_2, V_7\}, \{V_1, V_3, V_4, V_5, V_6\}).
\end{array} \tag{7}$$

(RLMP) invokes the following conditional independence relations:

$$\begin{array}{ll}
I(\{V_1\}, \emptyset, \{V_2, V_4, V_6, V_7, V_8, V_9\}), & I(\{V_2\}, \emptyset, \{V_1, V_3, V_5\}), \\
I(\{V_3\}, \{V_1\}, \{V_2, V_4, V_6, V_7, V_8, V_9\}), & I(\{V_4\}, \{V_2\}, \{V_1, V_3, V_5\}), \\
I(\{V_5\}, \{V_3\}, \{V_1, V_2, V_4, V_7, V_9\}), & I(\{V_6\}, \{V_4\}, \{V_1, V_2, V_3\}), \\
I(\{V_7\}, \{V_4\}, \{V_1, V_2, V_3, V_5\}), & I(\{V_8\}, \{V_6\}, \{V_1, V_2, V_3, V_4, V_5, V_7\}), \\
I(\{V_9\}, \{V_2, V_7\}, \{V_1, V_3, V_4, V_5, V_6\}) &
\end{array} \tag{8}$$

which, by Theorem 11, imply all the conditional independence relations in (7).

For the special case of graphs containing only bi-directed edges,[3] Kauermann (1996) provides a local Markov property for probability distributions obeying the composition axiom as follows:

$$\forall X \in V, \quad I(\{X\}, \emptyset, V \setminus (\{X\} \cup \text{sp}_G(X))). \tag{9}$$

3. Kauermann (1996) actually used undirected graphs with dashed edges which are Markov equivalent to graphs with only bi-directed edges (see Richardson, 2003, for discussions).

Since a graph containing only bi-directed edges is a special case of ADMGs without directed mixed cycles, the reduced local Markov property (RLMP) is applicable, and it turns out that (RLMP) reduces to (9) for graphs containing only bi-directed edges. Therefore (RLMP) includes the local Markov property given in Kauermann (1996) as a special case.

3.2 The Ordered Reduced Local Markov Property

The set of zero partial correlations corresponding to a conditional independence relation $I(X,Z,Y)$ is

$$\{\rho_{V_iV_j.Z} = 0 \mid V_i \in X, V_j \in Y\}.$$

Although (RLMP) gives only a linear number of conditional independence relations, the number of zero partial correlations may be larger than that invoked by (LMP,$\prec$) in some cases. For example, 12 conditional independence relations in (7) involve 37 zero partial correlations while 9 conditional independence relations in (8) involve 41 zero partial correlations. In this section, we will show an ordered local Markov property such that at most one zero partial correlation is invoked for each pair of variables.

Definition 12 (C-ordering) *Let G be an ADMG. A consistent ordering $\prec$ on the vertices of G is said to be a c-ordering if all the vertices in each c-component of G are consecutively ordered in $\prec$.*

For example, the ordering $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$ is a c-ordering on the vertices of G in Figure 3 (a). The following holds.

Proposition 13 *There exists a c-ordering on the vertices of G if G does not have directed mixed cycles.*

We can easily construct a c-ordering from the compressed graph of G. We introduce the following Markov property.

Definition 14 (The Ordered Reduced Local Markov Property (RLMP,$\prec_c$)) *Let G be an ADMG without directed mixed cycles and $\prec_c$ be a c-ordering on the vertices of G. A probability distribution P is said to satisfy the ordered reduced local Markov property for G with respect to $\prec_c$ if*

$$\text{(RLMP,}\prec_c\text{)} \qquad \forall X \in V,\ I(\{X\}, \mathrm{pa}_G(X), \mathrm{pre}_{G,\prec_c}(X) \setminus (\{X\} \cup \mathrm{pa}_G(X) \cup \mathrm{sp}_G(X))). \qquad (10)$$

The ordered reduced local Markov property states that *a variable is independent of its predecessors, excluding its spouses, in a c-ordering given its parents*. We now establish the equivalence of (GMP) and (RLMP,$\prec_c$).

Theorem 15 *If a probability distribution P satisfies the composition axiom and an ADMG G has no directed mixed cycles, then for a c-ordering $\prec_c$ on the vertices of G,*

$$\text{(GMP)} \Longleftrightarrow \text{(RLMP,}\prec_c\text{)}.$$

Proof: (GMP) $\Longrightarrow$ (RLMP,$\prec_c$)
The set $\mathrm{pre}_{G,\prec_c}(X)$ does not include any descendant of $\mathrm{dis}_G(X)$ since $\prec_c$ is a c-ordering. We have

$$\begin{aligned}
&\mathrm{pre}_{G,\prec_c}(X)\setminus(\{X\}\cup\mathrm{pa}_G(X)\cup\mathrm{sp}_G(X))\\
&=\mathrm{pre}_{G,\prec_c}(X)\setminus\Bigg(\Big(\{X\}\cup\mathrm{pa}_G(X)\cup\mathrm{sp}_G(X)\Big)\cup\Big(\mathrm{de}_G(\{X\}\cup\mathrm{sp}_G(X))\setminus(\{X\}\cup\mathrm{sp}_G(X))\Big)\Bigg)\\
&=\mathrm{pre}_{G,\prec_c}(X)\setminus\mathrm{f}(X,G)\\
&\subseteq V\setminus\mathrm{f}(X,G).
\end{aligned}$$

Hence, (RLMP,$\prec_c$) follows from (RLMP). ■

Proof: (RLMP,$\prec_c$) $\Longrightarrow$ (GMP)
We will show that (RLMP,$\prec_c$) $\Longrightarrow$ (LMP,$\prec_c$). Assume that a probability distribution P satisfies (RLMP,$\prec_c$). Let $\mathrm{g}(Y)=\mathrm{pre}_{G,\prec_c}(Y)\setminus(\{Y\}\cup\mathrm{pa}_G(Y)\cup\mathrm{sp}_G(Y))$. Consider the set of conditional independence relations invoked by (LMP,$\prec_c$) for each variable X given in (4) where A is maximal. By (10), for all Y in $\mathrm{dis}_{G_A}(X)$, we have

$$I(Y,\mathrm{pa}_G(Y),\mathrm{g}(Y)). \tag{11}$$

Let $S_1=\mathrm{pa}_G(\mathrm{dis}_{G_A}(X))\setminus\mathrm{pa}_G(Y)$ and $S_2=A\setminus(\mathrm{mb}(X,A)\cup\{X\})$. We have that

$$S_1\subseteq\mathrm{g}(Y).$$

Note that $S_2\setminus\mathrm{g}(Y)$ may be non-empty. Let $S_3=S_2\setminus\mathrm{g}(Y)$. It suffices to show that

$$I(Y,\mathrm{pa}_G(Y),S_3),$$

which implies $I(Y,\mathrm{pa}_G(Y),S_2)$ by composition. Then, the rest of the proof would be identical to that of Theorem 11.

We first characterize the vertices in S_3. We will show that

$$S_3=(\mathrm{pre}_{G,\prec_c}(X)\setminus\mathrm{pre}_{G,\prec_c}(Y))\setminus\mathrm{sp}_G(\mathrm{dis}_{G_A}(X)). \tag{12}$$

By Lemma 6, we have

$$S_2=\mathrm{pre}_{G,\prec_c}(X)\setminus\Big(\mathrm{de}_G(\mathrm{h}(X,A))\cup\mathrm{mb}(X,A)\cup\{X\}\Big).$$

Since $\prec_c$ is a c-ordering, no descendant of $\mathrm{dis}_G(X)$ will appear in A. Hence,

$$S_2=\mathrm{pre}_{G,\prec_c}(X)\setminus\Big(\mathrm{sp}_G(\mathrm{dis}_{G_A}(X))\cup\mathrm{pa}_G(\mathrm{dis}_{G_A}(X))\Big).$$

To identify some common elements of S_2 and $\mathrm{g}(Y)$, we will reformulate S_2 and $\mathrm{g}(Y)$ as follows.

$$\begin{aligned}
S_2&=\Big(B\setminus\mathrm{pa}_G(\mathrm{dis}_{G_A}(X))\Big)\cup\Big((\mathrm{dis}_G(X)\cap\mathrm{pre}_{G,\prec_c}(X))\setminus\mathrm{sp}_G(\mathrm{dis}_{G_A}(X))\Big),\\
\mathrm{g}(Y)&=\Big(B\setminus\mathrm{pa}_G(Y)\Big)\cup\Big((\mathrm{dis}_G(X)\cap\mathrm{pre}_{G,\prec_c}(Y))\setminus(\{Y\}\cup\mathrm{sp}_G(Y))\Big)
\end{aligned}$$

where $B = \mathrm{pre}_{G,\prec_c}(X) \setminus \mathrm{dis}_G(X)$. This can be verified by noting that $A_1 = A_2 \setminus (A_3 \cup A_4) = (A_{11} \setminus A_2) \cup (A_{12} \setminus A_3)$ if $A_1 = A_{11} \cup A_{12}, A_{11} \cap A_{12} = \emptyset, A_2 \subseteq A_{11}, A_3 \subseteq A_{12}$. From $\mathrm{pa}_G(Y) \subseteq \mathrm{pa}_G(\mathrm{dis}_{G_A}(X))$, it follows that $B \setminus \mathrm{pa}_G(\mathrm{dis}_{G_A}(X)) \subseteq B \setminus \mathrm{pa}_G(Y)$ and

$$\begin{aligned} S_3 =& S_2 \setminus \mathrm{g}(Y) \\ =& \Big((\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec_c}(X)) \setminus \mathrm{sp}_G(\mathrm{dis}_{G_A}(X)) \Big) \\ & \setminus \Big((\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec_c}(Y)) \setminus (\{Y\} \cup \mathrm{sp}_G(Y)) \Big). \end{aligned}$$

We can rewrite the first part of this expression as follows.

$$\begin{aligned} & (\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec_c}(X)) \setminus \mathrm{sp}_G(\mathrm{dis}_{G_A}(X)) \\ =& \Big((\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec_c}(Y)) \setminus \mathrm{sp}_G(\mathrm{dis}_{G_A}(X)) \Big) \\ & \cup \Big((\mathrm{pre}_{G,\prec_c}(X) \setminus \mathrm{pre}_{G,\prec_c}(Y)) \setminus \mathrm{sp}_G(\mathrm{dis}_{G_A}(X)) \Big). \end{aligned}$$

From $(\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec_c}(Y)) \setminus \mathrm{sp}_G(\mathrm{dis}_{G_A}(X)) \subseteq (\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec_c}(Y)) \setminus (\{Y\} \cup \mathrm{sp}_G(Y))$, (12) follows. Thus, the vertices in S_3 are those in the set $\mathrm{pre}_{G,\prec_c}(X) \setminus \mathrm{pre}_{G,\prec_c}(Y)$ and not in the set $\mathrm{sp}_G(\mathrm{dis}_{G_A}(X))$.

Now we are ready to prove $I(Y, \mathrm{pa}_G(Y), S_3)$. For any $Z \in S_3$, we have $Y \prec Z$ and $Z \notin \mathrm{sp}_G(Y)$. Hence,

$$\begin{array}{ll} I(\{Z\}, \mathrm{pa}_G(Z), \mathrm{g}(Z)), & \\ I(\{Z\}, \mathrm{pa}_G(Z), \{Y\} \cup (\mathrm{pa}_G(Y) \setminus \mathrm{pa}_G(Z))) & \text{by decomposition,} \\ I(\{Z\}, \mathrm{pa}_G(Z) \cup \mathrm{pa}_G(Y), \{Y\}) & \text{by weak union,} \\ I(\{Y\}, \mathrm{pa}_G(Y), \mathrm{pa}_G(Z) \setminus \mathrm{pa}_G(Y)) & \text{by } \mathrm{pa}_G(Z) \setminus \mathrm{pa}_G(Y)) \subseteq \mathrm{g}(Y), (11) \\ \quad \text{and decomposition,} & \\ I(\{Y\}, \mathrm{pa}_G(Y), \{Z\}) & \text{by contraction and decomposition.} \end{array}$$

Therefore, by composition, $I(Y, \mathrm{pa}_G(Y), S_3)$ holds. ■

(RLMP,$\prec_c$) invokes one zero partial correlation for each pair of nonadjacent variables. For example, for the ADMG G in Figure 3 (a) and a c-ordering $\prec_c$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$, (RLMP,$\prec_c$) invokes the following conditional independence relations:

$$\begin{array}{ll} I(\{V_2\}, \emptyset, \{V_1\}), & I(\{V_3\}, \{V_1\}, \{V_2\}), \\ I(\{V_4\}, \{V_2\}, \{V_1, V_3\}), & I(\{V_5\}, \{V_3\}, \{V_1, V_2, V_4\}), \\ I(\{V_6\}, \{V_4\}, \{V_1, V_2, V_3\}), & I(\{V_7\}, \{V_4\}, \{V_1, V_2, V_3, V_5\}), \\ I(\{V_8\}, \{V_6\}, \{V_1, V_2, V_3, V_4, V_5, V_7\}), & I(\{V_9\}, \{V_2, V_7\}, \{V_1, V_3, V_4, V_5, V_6\}) \end{array} \qquad (13)$$

which involve 25 zero partial correlations while (7) involve 37 zero partial correlations.

3.3 The Pairwise Markov Property

In this section, we give a pairwise Markov property which specifies conditional independence relations between pairs of variables and show that it is equivalent to the global Markov property. In previous sections, we focused on minimizing the number of zero partial correlations. We now take into account the size of the conditioning set Z in each zero partial correlation $\rho_{XY.Z}$. When the size of $\text{pa}_G(X)$ for a vertex X in (RLMP,$\prec_c$) is large, it might be advantageous to use a different conditioning set with smaller size (if the equivalence of the Markov properties still holds). Pearl and Meshkat (1999) introduced a pairwise Markov property for DAGs (without bi-directed edges) which may involve fewer conditioning variables and thus lead to more economical tests. The result can be easily generalized to ADMGs with no directed mixed cycles.

Let $d(X,Y)$ denote the shortest distance between two vertices X and Y, that is, the number of edges in the shortest path between X and Y. Two vertices X and Y are nonadjacent if X and Y are not connected by a directed nor a bi-directed edge.

Definition 16 (The Pairwise Markov Property (PMP,$\prec_c$)) *Let G be an ADMG without directed mixed cycles and $\prec_c$ be a c-ordering on the vertices of G. A probability distribution P is said to satisfy the pairwise Markov property for G with respect to $\prec_c$ if for any two nonadjacent vertices $V_i, V_j, V_j \prec_c V_i$*

$$\text{(PMP,}\prec_c\text{)} \qquad\qquad I(\{V_i\}, Z_{ij}, \{V_j\})$$

where Z_{ij} is any set of vertices such that Z_{ij} d-separates V_i from V_j and $\forall Z \in Z_{ij}, d(V_i, Z) < d(V_i, V_j)$.

Note that, in ADMGs with no directed mixed cycles, there always exists such a Z_{ij} for any two nonadjacent vertices. For example, the parent set of V_i always satisfies the condition for Z_{ij}. If the empty set d-separates V_i from V_j, then the empty set is defined to satisfy the condition for Z_{ij}. Therefore we can always choose a Z_{ij} with the smallest size, providing a more economical way to test zero partial correlations.

Theorem 17 *If a probability distribution P satisfies the composition axiom and an ADMG G has no directed mixed cycles, then*

$$\text{(GMP)} \Longleftrightarrow \text{(PMP,}\prec_c\text{)}.$$

Proof: Noting that two vertices X and Y are adjacent if $X \leftarrow Y$, $X \rightarrow Y$ or $X \leftrightarrow Y$, the proof of Theorem 1 by Pearl and Meshkat (1999) is directly applicable to ADMGs and it effectively proves that (RLMP,$\prec_c$) $\Longleftrightarrow$ (PMP,$\prec_c$). We do not reproduce the proof here. ■

As an example, for the ADMG G in Figure 3 (a) and a c-ordering $\prec_c$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$, the following conditional independence relations (for convenience, we combine the relations for each vertex that have the same conditioning set) can be given by (PMP,$\prec_c$):

$$\begin{array}{ll}
I(\{V_2\},\emptyset,\{V_1\}), & I(\{V_3\},\emptyset,\{V_2\}), \\
I(\{V_4\},\emptyset,\{V_3,V_1\}), & I(\{V_5\},\emptyset,\{V_4,V_2\}), \\
I(\{V_5\},\{V_3\},\{V_1\}), & I(\{V_6\},\emptyset,\{V_3,V_1\}), \\
I(\{V_6\},\{V_4\},\{V_2\}), & I(\{V_7\},\emptyset,\{V_5,V_3,V_1\}), \\
I(\{V_7\},\{V_4\},\{V_2\}), & I(\{V_8\},\{V_6\},\{V_7,V_5,V_4,V_2\}), \\
I(\{V_8\},\emptyset,\{V_3,V_1\}), & I(\{V_9\},\{V_2,V_7\},\{V_6,V_4\}), \\
I(\{V_9\},\emptyset,\{V_5,V_3,V_1\}) &
\end{array}$$

which involve the same number of zero partial correlations as (13) but involve smaller conditioning sets than those in (13).

3.4 Relation to Other Work

In this section, we contrast the class of ADMGs without directed mixed cycles to maximal ancestral graphs and chain graphs in terms of Markov properties.

3.4.1 MAXIMAL ANCESTRAL GRAPHS

It is easy to see that an ADMG without directed mixed cycles is a *maximal ancestral graph (MAG)* (Richardson and Spirtes, 2002). An ADMG is said to be *ancestral* if, for any edge $X \leftrightarrow Y$, X is not an ancestor of Y (and vice versa). Note that an edge $X \leftrightarrow Y$ and a directed path from X to Y (or Y to X) form a directed mixed cycle. Hence, an ADMG without directed mixed cycles is ancestral. An ancestral graph is said to be *maximal* if, for any pair of nonadjacent vertices X and Y, there exists a set $Z \subseteq V \setminus \{X,Y\}$ that d-separates X from Y. From Theorem 17, it follows that an ADMG without directed mixed cycles is maximal. On the other hand, there exist MAGs which have directed mixed cycles (see Figure 4). Thus, the class of ADMGs without directed mixed cycles is a strict subclass of MAGs.

Richardson and Spirtes (2002, p.979) showed the following pairwise Markov property for a MAG G:

$$I(\{V_i\}, \mathrm{an}_G(\{V_i,V_j\}) \setminus \{V_i,V_j\}, \{V_j\})$$

for any two nonadjacent vertices V_i and V_j. Richardson and Spirtes (2002) proved that this pairwise Markov property implies the global Markov property assuming a Gaussian parametrization. This does not trivially imply our results in Section 3.3 and our results cannot be considered as a special case of the results on MAGs. The two pairwise Markov properties involve two different forms of conditioning sets. The pairwise Markov property for MAGs involves considerably larger conditioning sets than our pairwise Markov property: the conditioning set includes all ancestors of V_i and V_j, which is undesirable for our purpose of using the zero partial correlations to test a model.

Also, it should be stressed that our results do not depend on a specific parameterization. We only require the composition axiom to be satisfied. In contrast, Richardson and Spirtes (2002) consider only Gaussian parameterizations. It requires further study whether the pairwise Markov property for MAGs can be generalized to the class of distributions satisfying the composition axiom.

In the next section, we consider general ADMGs and try to eliminate redundant conditional independence relations from (LMP,$\prec$). The class of MAGs is clearly a (strict) subclass of ADMGs. Hence, given a MAG, we have two options: either we use the result in the next section or the pairwise Markov property for MAGs. Although the pairwise Markov property for MAGs gives fewer zero partial correlations (one for each nonadjacent pair of vertices), it is possible that in some cases we are better off using the result in the next section (because of the cost incurred by the large conditioning sets in the pairwise Markov property for MAGs). An example of this situation will be given in the next section.

Richardson and Spirtes (2002) also proved that for a Gaussian distribution encoded by a MAG all the constraints on the distribution (that is, on the covariance matrix) are implied by the vanishing partial correlations given by the global Markov property. Hence, this also holds in a linear SEM represented by an ADMG without directed mixed cycles which is a special type of MAG.

3.4.2 CHAIN GRAPHS

The graph that results from replacing bi-directed edges with undirected edges in an ADMG without directed mixed cycles is a *chain graph*. The class of chain graphs has been studied extensively (see Lauritzen, 1996, for a review).

Some Markov properties have been proposed for chain graphs. The first Markov property for chain graphs has been proposed by Lauritzen and Wermuth (1989) and Frydenberg (1990). Andersson et al. (2001) have introduced another Markov property. These two Markov properties do not correspond to the Markov property for ADMGs. Let G be an ADMG without directed mixed cycles and G' be the chain graph obtained by replacing bi-directed edges with undirected edges. In general, the set of conditional independence relations given by the Markov property for G is not equivalent to that given by either of the two Markov properties for chain graphs. However, there are other Markov properties for chain graphs that correspond to the Markov property for ADMGs without directed mixed cycles (Cox and Wermuth, 1993; Wermuth and Cox, 2001, 2004).[4]

4. Markov Properties for General ADMGs

When an ADMG G has directed mixed cycles, (RLMP), (RLMP,$\prec_c$), and (PMP,$\prec_c$) are no longer equivalent to (GMP) while (LMP,$\prec$) still is. In this section, we show that the number of conditional independence relations given by (LMP,$\prec$) for an arbitrary ADMG that might have directed mixed cycles can still be reduced. We introduce a procedure to reduce (LMP,$\prec$). We then give an example to illustrate the procedure.

4.1 Reducing the Ordered Local Markov Property

First, we introduce a lemma that gives a condition by which a conditional independence relation renders another conditional independence relation redundant.

Lemma 18 *Given an ADMG G, a consistent ordering $\prec$ on the vertices of G and a vertex X, assume that a probability distribution P satisfies the global Markov property for $G_{\mathrm{pre}_{G,\prec}(X)\setminus\{X\}}$. Let $A = \mathrm{pre}_{G,\prec}(X)$ and A' be a maximal ancestral set with respect to $\mathrm{mb}(X,A')$ such that $X \in A' \subset A$, $A' \cap \mathrm{dis}_{G_A}(X) = \mathrm{dis}_{G_{A'}}(X)$ and $\mathrm{pa}_G(\mathrm{dis}_{G_A}(X) \setminus \mathrm{dis}_{G_{A'}}(X)) \subseteq \mathrm{mb}(X,A')$. Then,*

$$I(\{X\}, \mathrm{mb}(X,A), A \setminus (\mathrm{mb}(X,A) \cup \{X\})) \tag{14}$$

implies

$$I(\{X\}, \mathrm{mb}(X,A'), A' \setminus (\mathrm{mb}(X,A') \cup \{X\})).$$

We define $\mathrm{rd}_{G,\prec}(X)$ to be the set of all A' satisfying this condition.

Proof: First, we show the relationships among $A, \mathrm{dis}_{G_A}(X)$, $\mathrm{mb}(X,A)$ and $A', \mathrm{dis}_{G_{A'}}(X)$, $\mathrm{mb}(X,A')$. By Lemma 6, we have

$$A' = A \setminus \mathrm{de}_{G_A}(\mathrm{h}(X,A')) \tag{15}$$

where

$$\mathrm{h}(X,A') \equiv \mathrm{sp}_{G_A}\Big(\mathrm{dis}_{G_{A'}}(X)\Big) \setminus \Big(\{X\} \cup \mathrm{mb}(X,A')\Big).$$

4. In their terminology, ADMGs without directed mixed cycles correspond to chain graphs with dashed arrows and dashed edges.

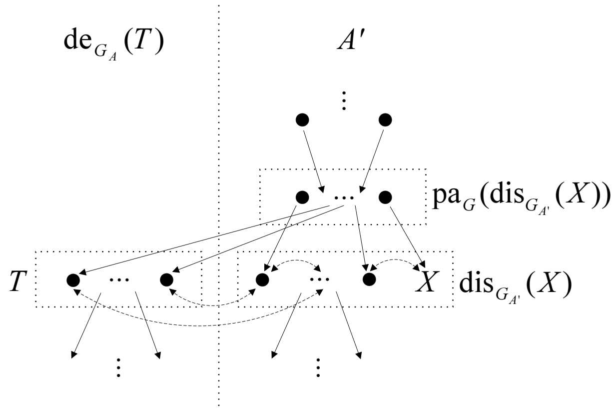


Figure 5: The relationship between A and A' that satisfy the conditions in Lemma 18. The induced subgraph G_A is shown. The vertices of G_A are decomposed into two disjoint subsets $\mathrm{de}_{G_A}(T)$ and A'.

$\mathrm{dis}_{G_{A'}}(X)$ and $\mathrm{h}(X,A')$ are subsets of $\mathrm{dis}_{G_A}(X)$. Since $\mathrm{dis}_{G_{A'}}(X) \subseteq \{X\} \cup \mathrm{mb}(X,A')$ (by the definition of the Markov blanket), $\mathrm{dis}_{G_{A'}}(X) \cap \mathrm{h}(X,A') = \emptyset$. Thus, we can decompose the set $\mathrm{dis}_{G_A}(X)$ into 3 disjoint subsets as follows.

$$\mathrm{dis}_{G_A}(X) = \mathrm{dis}_{G_{A'}}(X) \cup \mathrm{h}(X,A') \cup B \tag{16}$$

where

$$B \equiv \mathrm{dis}_{G_A}(X) \setminus \Big(\mathrm{dis}_{G_{A'}}(X) \cup \mathrm{h}(X,A')\Big).$$

We have

$$\begin{aligned} A' \cap \mathrm{dis}_{G_A}(X) &= A' \cap \Big(\mathrm{dis}_{G_{A'}}(X) \cup \mathrm{h}(X,A') \cup B\Big) \\ &= \mathrm{dis}_{G_{A'}}(X) \cup B \end{aligned}$$

since $\mathrm{dis}_{G_{A'}}(X) \subseteq A', B \subseteq A'$ and $A' \cap \mathrm{h}(X,A') = \emptyset$. From the assumption in Lemma 18 that $A' \cap \mathrm{dis}_{G_A}(X) = \mathrm{dis}_{G_{A'}}(X)$, it follows that $B = \emptyset$. Thus, from (16), we have

$$\mathrm{dis}_{G_A}(X) \setminus \mathrm{dis}_{G_{A'}}(X) = \mathrm{h}(X,A'). \tag{17}$$

Let $T = \mathrm{dis}_{G_A}(X) \setminus \mathrm{dis}_{G_{A'}}(X) = \mathrm{h}(X,A')$. Then,

$$\begin{aligned} \mathrm{mb}(X,A) &= \mathrm{mb}(X,A') \cup T \cup \mathrm{pa}_G(T) \\ &= \mathrm{mb}(X,A') \cup T \end{aligned} \tag{18}$$

since $\mathrm{pa}_G(T) \subseteq \mathrm{mb}(X,A')$ by our assumption. Thus A decomposes into

$$A = A' \cup \mathrm{de}_{G_A}(T) \tag{19}$$

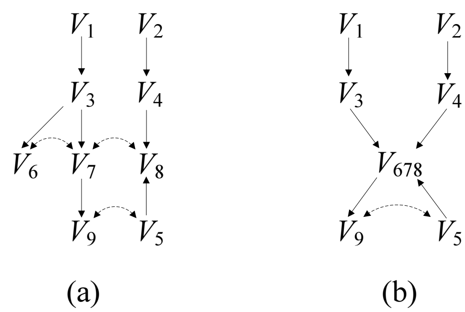


Figure 6: (a) An ADMG with directed mixed cycles (b) Illustration of the procedure **GetOrdering**. The modified graph after the first step is shown.

since $\mathrm{de}_{G_A}(T) \subseteq A$ and (15).

The key relationships among $A, \mathrm{dis}_{G_A}(X), \mathrm{mb}(X,A)$ and $A', \mathrm{dis}_{G_{A'}}(X), \mathrm{mb}(X,A')$ are given by (17)–(19). Figure 5 shows these relationships. We are now ready to prove that $I(\{X\}, \mathrm{mb}(X,A'), A' \setminus (\mathrm{mb}(X,A') \cup \{X\}))$ can be derived from $I(\{X\}, \mathrm{mb}(X,A), A \setminus (\mathrm{mb}(X,A) \cup \{X\}))$. From (18) and (19), it follows that

$$A \setminus (\mathrm{mb}(X,A) \cup \{X\}) = (A' \cup \mathrm{de}_{G_A}(T)) \setminus (\mathrm{mb}(X,A') \cup \{X\} \cup T).$$

Since $A' \cap \mathrm{de}_{G_A}(T) = \emptyset, (\mathrm{mb}(X,A') \cup \{X\}) \cap T = \emptyset, \mathrm{mb}(X,A') \cup \{X\} \subseteq A'$ and $T \subseteq \mathrm{de}_{G_A}(T)$, we have

$$A \setminus (\mathrm{mb}(X,A) \cup \{X\}) = \Big(A' \setminus (\mathrm{mb}(X,A') \cup \{X\})\Big) \cup \Big(\mathrm{de}_{G_A}(T) \setminus T\Big). \tag{20}$$

Plugging (18) and (20) into (14), we get

$$I\Big(\{X\}, \mathrm{mb}(X,A') \cup T, \Big(A' \setminus (\mathrm{mb}(X,A') \cup \{X\})\Big) \cup \Big(\mathrm{de}_{G_A}(T) \setminus T\Big)\Big).$$

From the decomposition axiom, it follows that

$$I(\{X\}, \mathrm{mb}(X,A') \cup T, A' \setminus (\mathrm{mb}(X,A') \cup \{X\})). \tag{21}$$

The last step is to remove T from the conditioning set to obtain $I(\{X\}, \mathrm{mb}(X,A'), A' \setminus (\mathrm{mb}(X,A') \cup \{X\}))$. We claim that

$$I(T, \mathrm{mb}(X,A'), A' \setminus (\mathrm{mb}(X,A') \cup \{X\})). \tag{22}$$

We first argue that T is d-separated from $A' \setminus (\mathrm{mb}(X,A') \cup \{X\})$ given $\mathrm{mb}(X,A')$. Consider a vertex $t \in T$ and a vertex $\alpha \in A' \setminus (\mathrm{mb}(X,A') \cup \{X\})$. Note that for any bi-directed edge $t \leftrightarrow \beta$ in G_A, β is either in T or $\mathrm{dis}_{G_{A'}}(X)$. There are only four possible cases for any path in G_A from t to α.

1. $t \leftarrow \gamma \cdots \alpha$
2. $t \rightarrow \cdots \rightarrow \gamma \leftarrow\!\!* \cdots \alpha$

3. $t \leftrightarrow\leftrightarrow \cdots \leftrightarrow \delta \leftarrow \gamma \cdots \alpha$

4. $t \leftrightarrow\leftrightarrow \cdots \leftrightarrow \delta \rightarrow \cdots \rightarrow \gamma \leftarrow\!\!* \cdots \alpha$

In case 1, $\gamma \in \text{mb}(X, A')$ since $\text{pa}_G(T) \subseteq \text{mb}(X, A')$. Thus, the path is not d-connecting. In case 2, γ is a descendant of t. Since $\text{mb}(X, A')$ does not contain any descendant of t, the path is not d-connecting. Case 3 is similar to case 1, but there are one or more bi-directed edges after t. δ is either in T or $\text{dis}_{G_{A'}}(X)$. It follows that $\gamma \in \text{mb}(X, A')$, so the path is not d-connecting. Case 4 is similar to case 2, but there are one or more bi-directed edges after t. If δ is in T, the argument for case 2 can be applied. If δ is in $\text{dis}_{G_{A'}}(X)$, then $\delta \in \text{mb}(X, A')$, which implies that the path is not d-connecting. This establishes that T is d-separated from $A' \setminus (\text{mb}(X, A') \cup \{X\})$ given $\text{mb}(X, A')$. By the assumption that P satisfies the global Markov property for $G_{\text{pre}_{G,\prec}(X) \setminus \{X\}}$, (22) holds. Finally, from (21),(22) and the contraction axiom, it follows that $I(\{X\}, \text{mb}(X, A'), A' \setminus (\text{mb}(X, A') \cup \{X\}))$. ■

For example, consider the ADMG G in Figure 2 and a consistent ordering $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7$. Assume that the global Markov property for $G_{\text{pre}_{G,\prec}(V_6)}$ is satisfied. Let $A = \{V_1, V_2, V_3, V_4, V_5, V_6, V_7\}$ and $A' = \{V_1, V_2, V_3, V_4, V_6, V_7\}$. Then, $\text{dis}_{G_A}(V_7) = \{V_5, V_6, V_7\}$, $\text{dis}_{G_{A'}}(V_7) = \{V_6, V_7\}$, $A' \cap \text{dis}_{G_A}(V_7) = \{V_6, V_7\} = \text{dis}_{G_{A'}}(V_7)$ and $\text{pa}_G(\text{dis}_{G_A}(V_7) \setminus \text{dis}_{G_{A'}}(V_7)) = \{V_3\} \subseteq \{V_3, V_4, V_6\} = \text{mb}(V_7, A')$. Thus, $I(\{V_7\}, \{V_3, V_4, V_6\}, \{V_1, V_2\})$ follows from $I(\{V_7\}, \{V_3, V_4, V_5, V_6\}, \{V_1, V_2\})$. Note that in the proof of Lemma 18, the composition axiom is not used. Thus, Lemma 18 can be used to reduce the ordered local Markov property for ADMGs associated with an arbitrary probability distribution. Also, note that the condition that P satisfies the global Markov property for $G_{\text{pre}_{G,\prec}(X) \setminus \{X\}}$ is always satisfied in a recursive application of this lemma in Theorem 21.

We now introduce a key concept in eliminating redundant conditional independence relations from (LMP,$\prec$).

Definition 19 (C-ordered Vertex) *Given a consistent ordering $\prec$ on the vertices of an ADMG G, a vertex X is said to be c-ordered in $\prec$ if*

1. *all vertices in $\text{dis}_G(X) \cap \text{pre}_{G,\prec}(X)$ are consecutive in $\prec$ and*

2. *for any two vertices Y and Z in $\text{dis}_G(X) \cap \text{pre}_{G,\prec}(X)$, there is no directed edge between Y and Z.*

If no bi-directed edge is connected to X, then X is defined to be c-ordered. For example, consider the ADMG G in Figure 6 (a). $\prec$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$ is a consistent ordering on the vertices of G. $V_1, V_2, \ldots, V_8$ are c-ordered in $\prec$ but V_9 is not since V_5 and V_9 are not consecutive in $\prec$.

The key observation, which will be proved, is that c-ordered vertices contribute to eliminating many redundant conditional independence relations invoked by the ordered local Markov property (LMP,$\prec$). We provide two procedures. The first procedure **ReduceMarkov** in Figure 7 constructs a list of conditional independence relations in which some redundant conditional independence relations from (LMP,$\prec$) are not included (all the conditional independence relations identified by Lemma 18 are not included). **ReduceMarkov** takes as input a fixed ordering $\prec$. The second procedure **GetOrdering** in Figure 9 gives a good ordering that might have many c-ordered vertices.

We first describe the procedure **ReduceMarkov**. Given an ADMG G and a consistent ordering $\prec$, **ReduceMarkov** gives a set of conditional independence relations which will be shown to be

procedure ReduceMarkov
INPUT: An ADMG G and a consistent ordering $\prec$ on the vertices of G
OUTPUT: A set of conditional independence relations S
$S \leftarrow \emptyset$
for $i = 1, \ldots, n$ **do**
  $I_i \leftarrow \emptyset$
  if V_i is c-ordered in $\prec$ **then**
    for nonadjacent $V_j \prec V_i$ **do**
      $I_i \leftarrow I_i \cup I(\{V_i\}, Z_{ij}, \{V_j\})$ where Z_{ij} is any set of vertices such that Z_{ij} d-separates V_i from V_j and $\forall Z \in Z_{ij}, d(V_i, Z) < d(V_i, V_j)$
    end for
  else
    for all maximal ancestral sets A with respect to $\mathrm{mb}(V_i, A)$ such that $V_i \in A \subseteq \mathrm{pre}_{G,\prec}(V_i)$, $A \notin \mathrm{rd}_{G,\prec}(V_i)$ **do**
      $I_i \leftarrow I_i \cup I(\{V_i\}, \mathrm{mb}(V_i, A), A \setminus (\mathrm{mb}(V_i, A) \cup \{V_i\}))$
    end for
  end if
  $S \leftarrow S \cup I_i$
end for

Figure 7: A procedure to generate a reduced set of conditional independence relations for an ADMG G and a consistent ordering $\prec$

equivalent to the global Markov property for G. For each vertex V_i, **ReduceMarkov** generates a set of conditional independence relations. If V_i is c-ordered, the relations that correspond to the pairwise Markov property are generated. Otherwise, the relations that correspond to the ordered local Markov property are generated, and Lemma 18 is used to remove some redundant relations. The output $S = \textbf{ReduceMarkov}(G, \prec)$ can be described as follows:

$$S = \bigcup_{X:X \text{ is c-ordered in } \prec} \Big(\bigcup_{Y:Y \prec X} I\big(\{X\}, Z_{XY}, \{Y\}\big)\Big) \cup \bigcup_{X:X \text{ is not c-ordered in } \prec} \Bigg(\bigcup_{\substack{\text{all maximal sets } A \\ \text{with respect to } \mathrm{mb}(X,A): \\ X \in A \subseteq \mathrm{pre}_{G,\prec}(X), \\ A \notin \mathrm{rd}_{G,\prec}(X)}} I\big(\{X\}, \mathrm{mb}(X,A), A \setminus (\mathrm{mb}(X,A) \cup \{X\})\big)\Bigg) \tag{23}$$

where Z_{XY} is any set of vertices such that Z_{XY} d-separates X from Y and $\forall Z \in Z_{XY}$, $d(X,Z) < d(X,Y)$.

If a vertex X is c-ordered, $O(n)$ conditional independence relations (or zero partial correlations) are added to S. Otherwise, $O(2^n)$ conditional independence relations may be added to S and $O(n2^n)$ zero partial correlations may be invoked. Furthermore, a c-ordered vertex typically involves a smaller conditioning set. $I(\{X\}, Z_{XY}, \{Y\})$ has the conditioning set $|Z_{XY}| \leq |\mathrm{pa}_G(X)|$ while $I(\{X\}, \mathrm{mb}(X,A), A \setminus (\mathrm{mb}(X,A) \cup \{X\}))$ has the conditioning set $|\mathrm{mb}(X,A)| \geq |\mathrm{pa}_G(X)|$.

We now prove that the conditional independence relations produced by **ReduceMarkov** can derive all the conditional independence relations invoked by the global Markov property.

Definition 20 (S-Markov Property (S-MP,$\prec$)) *Let G be an ADMG and $\prec$ be a consistent ordering on the vertices of G. Let S be the set of conditional independence relations given by **ReduceMarkov**(G,$\prec$). A probability distribution P is said to satisfy the S-Markov property for G with respect to $\prec$, if*

(S-MP,$\prec$) *P satisfies all the conditional independence relations in S.*

Theorem 21 *Let G be an ADMG and $\prec$ be a consistent ordering on the vertices of G. Let S be the set of conditional independence relations given by **ReduceMarkov**(G,$\prec$). If a probability distribution P satisfies the composition axiom, then*

$$\text{(GMP)} \Longleftrightarrow (S\text{-MP},\prec).$$

Proof: (GMP) $\Longrightarrow$ (S-MP,$\prec$) since every conditional independence relation in (S-MP,$\prec$) corresponds to a valid d-separation. We show (S-MP,$\prec$) $\Longrightarrow$ (GMP). Without any loss of generality, let $\prec: V_1 \prec \ldots \prec V_n$. The proof is by induction on the sequence of ordered vertices. Suppose that (S-MP,$\prec$) $\Longrightarrow$ (GMP) holds for $V_1,\ldots V_{i-1}$. Let $S_{i-1} = I_1 \cup \ldots \cup I_{i-1}$. Then, by the induction hypothesis, S_{i-1} contains all the conditional independence relations invoked by (LMP,$\prec$) for $V_1,\ldots V_{i-1}$. If V_i is not c-ordered, $I_i = I(\{V_i\}, \mathrm{mb}(V_i,A), A\setminus(\mathrm{mb}(V_i,A)\cup\{V_i\}))$ for all maximal ancestral sets A such that $V_i \in A \subseteq \mathrm{pre}_{G,\prec}(V_i)$, $A \notin \mathrm{rd}_{G,\prec}(V_i)$. The conditional independence relations invoked by (LMP,$\prec$) with respect to V_i and any $A \in \mathrm{rd}_{G,\prec}(V_i)$ can be derived from other conditional independence relations by Lemma 18. Thus, $S_i = S_{i-1} \cup I_i$ contains all the conditional independence relations invoked by (LMP,$\prec$) for $V_1,\ldots V_i$, which implies (GMP). If V_i is c-ordered, applying the arguments in the proof of (GMP) $\Longleftrightarrow$ (PMP,$\prec_c$), we have

$$I(\{V_i\}, \mathrm{pa}_G(V_i), \mathrm{pre}_{G,\prec}(V_i)\setminus(\{V_i\}\cup \mathrm{pa}_G(V_i)\cup \mathrm{sp}_G(V_i))).$$

By the induction hypothesis and the definition of a c-ordered vertex, we have for all $V_j \in \mathrm{dis}_G(V_i)\cap \mathrm{pre}_{G,\prec}(V_i)$

$$I(\{V_j\}, \mathrm{pa}_G(V_j), \mathrm{pre}_{G,\prec}(V_j)\setminus(\{V_j\}\cup \mathrm{pa}_G(V_j)\cup \mathrm{sp}_G(V_j))).$$

By the arguments in the proof of (GMP) $\Longleftrightarrow$ (RLMP,$\prec_c$), we have for all maximal ancestral sets A such that $V_i \in A \subseteq \mathrm{pre}_{G,\prec}(V_i)$

$$I(\{V_i\}, \mathrm{mb}(V_i,A), A\setminus(\mathrm{mb}(V_i,A)\cup\{V_i\})).$$

Therefore, $S_i = S_{i-1} \cup I_i$ derives all the conditional independence relations invoked by (GMP). ■

As we have seen earlier, the number of zero partial correlations critically depends on the number of c-ordered vertices in a given ordering. This motivates us to find the ordering with the most c-ordered vertices. An obvious way of finding this ordering is to explore the space of all the consistent orderings. However, this exhaustive search may become infeasible as the number of vertices grows. We propose a greedy algorithm to get an ordering that has a large number of c-ordered vertices. The basic idea is to first find a large c-component in which many vertices can be c-ordered and place the vertices consecutively in the ordering, then repeating this until we cannot find a set of vertices that can be c-ordered. To describe the algorithm, we define the following notion, which identifies the largest subset of a c-component that can be c-ordered.

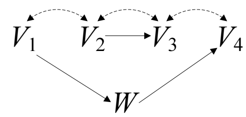


Figure 8: The c-component $\{V_1, V_2, V_3, V_4\}$ has the root set $\{V_1, V_2\}$

Definition 22 (Root Set) *The root set of a c-component C, denoted $\mathrm{rt}(C)$ is defined to be the set $\{V_i \in C \mid$ there is no $V_j \in C$ such that a directed path $V_j \to \ldots \to V_i$ exists in $G\}$.*

For example, the c-component $\{V_1, V_2, V_3, V_4\}$ in Figure 8 has the root set $\{V_1, V_2\}$. V_3 and V_4 are not in the root set since there are paths $V_2 \to V_3$ and $V_1 \to W \to V_4$. The root set has the following properties.

Proposition 23 *Let $\prec$ be a consistent ordering on the vertices of an ADMG G and C be a c-component of G. If the vertices in $\mathrm{rt}(C)$ are consecutive in $\prec$, then all the vertices in $\mathrm{rt}(C)$ are c-ordered in $\prec$.*

Proof: Assume that the vertices in $\mathrm{rt}(C)$ are consecutive in $\prec$. Then, for $X \in \mathrm{rt}(C)$, $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X) \subseteq \mathrm{rt}(C)$. Thus, there is no directed edge between any two vertices in $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$. ■

Proposition 24 *Let $\prec$ be a consistent ordering on the vertices of an ADMG G and C be a c-component of G. If a vertex X in C is c-ordered in $\prec$, then $X \in \mathrm{rt}(C)$.*

Proof: Assume that X is c-ordered in $\prec$. Suppose for a contradiction that $X \notin \mathrm{rt}(C)$. Then, there exists an ancestor Y of X in C. If there exists a vertex Z such that $Z \notin C$, $Y \to \cdots \to Z \to \cdots \to X$. Then, the first condition of a c-ordered vertex is violated. Otherwise, the second condition is violated. ■

Proposition 23 and 24 imply that the root set of a c-component is the largest subset of the c-component that can be c-ordered in a consistent ordering. If G does not have directed mixed cycles, $\mathrm{rt}(C) = C$ for every c-component C.

The procedure **GetOrdering** in Figure 9 is our proposed greedy algorithm that generates a good consistent ordering for G. In Step 1, it searches for the largest root set M and then merges all the vertices in M to one vertex V_M modifying edges accordingly. Then, it repeats the same operation for the modified graph until there is no root set that contains more than one vertex. Since the vertices in a root set are merged at each iteration, the modified graph is acyclic as otherwise there would be a directed path between two vertices in the root set, which contradicts the condition of a root set. After Step 1, we can easily obtain a consistent ordering for the original graph from the modified graph.

4.2 An Example

We show the application of the procedures **ReduceMarkov** and **GetOrdering** by considering the ADMG G in Figure 6 (a). First, we apply **GetOrdering** to get a consistent ordering on the vertices V of G. In Step 1, we first look for the largest root set. The c-component $\{V_6, V_7, V_8\}$ has the largest

procedure GetOrdering

INPUT: An ADMG G
OUTPUT: A consistent ordering $\prec$ on V
Step 1:
$G' \leftarrow G$ (V' is the set of vertices of G)
while (there is a c-component C of G' such that $|\mathrm{rt}(C)| > 1$) **do**
  $M \leftarrow \emptyset$
  for each c-component C of G' **do**
    if $|\mathrm{rt}(C)| > |M|$ **then**
      $M \leftarrow \mathrm{rt}(C)$
    end if
  end for
  Add a vertex V_M to $G'_{V' \setminus M}$
  Draw an edge $V_M \leftarrow X$ (respectively $V_M \rightarrow X$, $V_M \leftrightarrow X$) if there is
  $Y \leftarrow X$ (respectively $Y \rightarrow X$, $Y \leftrightarrow X$) in G' such that $Y \in M, X \in V' \setminus M$
  Let G' be the resulting graph
end while
Step 2:
Let $\prec'$ be any consistent ordering on V'. Construct a consistent ordering $\prec$ from $\prec'$ by replacing each $V_S \in V' \setminus V$ with the vertices in S (the ordering of the vertices in S is arbitrary)

Figure 9: A greedy algorithm to generate a good consistent ordering on the vertices of an ADMG G

root set $\{V_6, V_7, V_8\}$. Then, the vertices in $\{V_6, V_7, V_8\}$ are merged into a vertex V_{678}. Figure 6 (b) shows the modified graph G' after the first iteration of the while loop. In the next iteration, we find that every c-component has the root set of size 1. Note that for $C = \{V_5, V_9\}$, $\mathrm{rt}(C) = \{V_5, V_9\}$ in G but $\mathrm{rt}(C) = \{V_5\}$ in G'. Thus, Step 1 ends. In Step 2, from G' in Figure 6 (b), we can obtain an ordering $\prec'$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_{678} \prec V_9$. This is converted to a consistent ordering $\prec$: $V_1 \prec V_2 \prec V_3 \prec V_4 \prec V_5 \prec V_6 \prec V_7 \prec V_8 \prec V_9$ for G.

With the ordering $\prec$, we now apply **ReduceMarkov** to obtain a set of conditional independence relations that can derive those invoked by the global Markov property. It is easy to see that the vertices $V_1, \ldots, V_8$ are c-ordered in $\prec$. Thus, the following conditional independence relations corresponding to the pairwise Markov property are added to the set S (initially empty).

$$\begin{aligned}
&I(\{V_2\}, \emptyset, \{V_1\}), && I(\{V_3\}, \emptyset, \{V_2\}),\\
&I(\{V_4\}, \emptyset, \{V_3, V_1\}), && I(\{V_5\}, \emptyset, \{V_4, V_3, V_2, V_1\}),\\
&I(\{V_6\}, \emptyset, \{V_5, V_4, V_2\}), && I(\{V_6\}, \{V_3\}, \{V_1\}),\\
&I(\{V_7\}, \emptyset, \{V_5, V_4, V_2\}), && I(\{V_7\}, \{V_3\}, \{V_1\}),\\
&I(\{V_8\}, \emptyset, \{V_6, V_3, V_1\}), && I(\{V_8\}, \{V_4\}, \{V_2\}).
\end{aligned} \tag{24}$$

V_9 is not c-ordered in $\prec$ since V_5 is not adjacent in $\prec$. Thus, we use the ordered local Markov property (LMP,$\prec$) for V_9. The maximal ancestral sets that we need to consider are

$$A_1 = \mathrm{an}_G(\{V_6,V_8,V_9\}) = \{V_1,V_2,V_3,V_4,V_5,V_6,V_7,V_8,V_9\} \text{ and}$$
$$A_2 = \mathrm{an}_G(\{V_4,V_6,V_9\}) = \{V_1,V_2,V_3,V_4,V_6,V_7,V_9\}.$$

The corresponding conditional independence relations are

$$I(\{V_9\},\{V_7,V_5\},\{V_8,V_6,V_4,V_3,V_2,V_1\}), \tag{25}$$
$$I(\{V_9\},\{V_7\},\{V_6,V_4,V_3,V_2,V_1\}). \tag{26}$$

However, it turns out that $A_2 \in \mathrm{rd}_{G,\prec}(V_9)$ and (26) is not added to S. We check the condition of Lemma 18. The global Markov property for $G_{\mathrm{pre}_{G,\prec}(V_8)}$ is satisfied by (24). Also,

$$\begin{aligned}
\mathrm{dis}_{G_{A_1}}(V_9) &= \{V_5,V_9\},\\
\mathrm{dis}_{G_{A_2}}(V_9) &= \{V_9\},\\
A_2 \cap \mathrm{dis}_{G_{A_1}}(V_9) &= \{V_9\} = \mathrm{dis}_{G_{A_2}}(V_9),\\
\mathrm{pa}_G(\mathrm{dis}_{G_{A_1}}(V_9)\setminus \mathrm{dis}_{G_{A_2}}(V_9)) &= \emptyset \subseteq \{V_7\} = \mathrm{mb}(V_9,A_2).
\end{aligned}$$

Therefore, the condition of Lemma 18 is satisfied and it follows that (26) is redundant. To see how much we reduced the testing requirements, the conditional independence relations invoked by (LMP,$\prec$) are shown below.

$$\begin{array}{ll}
I(\{V_2\},\emptyset,\{V_1\}), & I(\{V_3\},\{V_1\},\{V_2\}),\\
I(\{V_4\},\{V_2\},\{V_3,V_1\}), & I(\{V_5\},\emptyset,\{V_4,V_3,V_2,V_1\}),\\
I(\{V_6\},\{V_3\},\{V_5,V_4,V_2,V_1\}), & I(\{V_7\},\{V_3\},\{V_5,V_4,V_2,V_1\}),\\
I(\{V_7\},\{V_6,V_3\},\{V_5,V_4,V_2,V_1\}), & I(\{V_8\},\{V_5,V_4\},\{V_6,V_3,V_2,V_1\}),\\
I(\{V_8\},\{V_7,V_5,V_4,V_3\},\{V_2,V_1\}), & I(\{V_8\},\{V_7,V_6,V_5,V_4,V_3\},\{V_2,V_1\}),\\
I(\{V_9\},\{V_7\},\{V_6,V_4,V_3,V_2,V_1\}), & I(\{V_9\},\{V_7,V_5\},\{V_8,V_6,V_4,V_3,V_2,V_1\}).
\end{array} \tag{27}$$

S invokes 26 zero partial correlations while (LMP,$\prec$) invokes 39. Also, S involves much smaller conditioning sets. We have at most one vertex in each conditioning set in (24) and two vertices in (25) while 23 zero partial correlations in (27) involve more than 2 vertices in the conditioning set.

The ADMG G in this example turns out to be a MAG. As we discussed in Section 3.4.1, we have two options: either we use the constraints in (24) and (25) or the constraints given by the pairwise Markov property for MAGs. In this example, both sets of constraints involve the same number of zero partial correlations. However, the pairwise Markov property for MAGs involves much larger conditioning sets. For example, the pairwise Markov property for MAGs gives the following conditional independence relation for the pair V_6 and V_8: $I(\{V_8\},\{V_5,V_4,V_3,V_2,V_1\},\{V_6\})$. Our method uses an empty set as the conditioning set for the pair. Hence, in this example, we are better off using the constraints in (24) and (25).

4.3 Comparison of (LMP,$\prec$) and (S-MP,$\prec$)

From (23), it is clear that (S-MP,$\prec$) invokes fewer conditional independence relations than (LMP,$\prec$) if there are c-ordered vertices in $\prec$. But how much more economical is (S-MP,$\prec$) than (LMP,$\prec$) and for what type of graphs is the reduction large?

For simplicity, we will compare the number of conditional independence relations rather than zero partial correlations and ignore the reduction done by Lemma 18. For now assume

$$S = \bigcup_{X: X \text{ is c-ordered in } \prec} I(\{X\}, \mathrm{pa}_G(X), \mathrm{pre}_{G,\prec}(X) \setminus (\{X\} \cup \mathrm{pa}_G(X) \cup \mathrm{sp}_G(X))) \bigcup$$
$$\bigcup_{X: X \text{ is not c-ordered in } \prec} \Bigg(\bigcup_{\substack{\text{all maximal sets } A \\ \text{with respect to } \mathrm{mb}(X,A): \\ X \in A \subseteq \mathrm{pre}_{G,\prec}(X)}} I\big(\{X\}, \mathrm{mb}(X,A), A \setminus (\mathrm{mb}(X,A) \cup \{X\})\big)\Bigg).$$

Let $\mathrm{M}(X,\prec)$ be the number of different Markov blankets of a vertex X, that is, $\mathrm{M}(X,\prec) = \Big|\{\mathrm{dis}_{G_A}(X) \mid A \text{ is an ancestral set such that } X \in A \subseteq \mathrm{pre}_{G,\prec}(X)\}\Big|$, and $\mathrm{C}(\prec)$ be the set of vertices that are c-ordered in $\prec$. Then, (LMP,$\prec$) lists $\sum_{X \in V} \mathrm{M}(X,\prec)$ conditional independence relations and (S-MP,$\prec$) lists $|\mathrm{C}(\prec)| + \sum_{X \notin \mathrm{C}(\prec)} \mathrm{M}(X,\prec)$ conditional independence relations. Hence, the difference in the number of conditional independence relations between (LMP,$\prec$) and (S-MP,$\prec$) is

$$\sum_{X \in \mathrm{C}(\prec)} \Big(\mathrm{M}(X,\prec) - 1\Big).$$

This difference is large when $|\mathrm{C}(\prec)|$ or $\mathrm{M}(X,\prec)$ for each X is large.

The size of $\mathrm{C}(\prec)$ depends on the number of directed mixed cycles. From Definition 19, it follows that $\mathrm{C}(\prec)$ is large if there are a small number of directed mixed cycles. Note that a directed mixed cycle such as that in Figure 4 induces the violation of the first condition in Definition 19 and a directed mixed cycle of the form $\alpha \rightleftarrows \beta$ induces the violation of the second condition in Definition 19.

$\mathrm{M}(X,\prec)$ depends on the structure of $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$. We will reformulate $\mathrm{M}(X,\prec)$ to show the properties that affect $\mathrm{M}(X,\prec)$. Let $G_{\leftrightarrow,\mathrm{dis}}(X,\prec) = (V',E')$ where $V' = \mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$ and $E' = \{V_i \leftrightarrow V_j \mid V_i \leftrightarrow V_j \text{ in } G_{V'}\}$. For example, for an ADMG G in Figure 8 and an ordering $V_1 \prec V_2 \prec V_3 \prec V_4$, $G_{\leftrightarrow,\mathrm{dis}}(V_3,\prec)$ is $V_1 \leftrightarrow V_2 \leftrightarrow V_3$. Let $G_{\leftrightarrow,\mathrm{dis}}(X,\prec)_S$ be the induced subgraph of $G_{\leftrightarrow,\mathrm{dis}}(X,\prec)$ on a set $S \subseteq \mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$. Then, $\mathrm{M}(X,\prec) = \Big|\{S \mid S \subseteq \mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$ such that $G_{\leftrightarrow,\mathrm{dis}}(X,\prec)_S$ is a *connected component* of $G_{\leftrightarrow,\mathrm{dis}}(X,\prec)_{S \,\cup\, (\mathrm{an}_G(S) \,\cap\, \mathrm{dis}_G(X) \,\cap\, \mathrm{pre}_{G,\prec}(X))}\}\Big|$ that is, $\mathrm{M}(X,\prec)$ corresponds to a set of subsets S of $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$ satisfying two conditions: (i) $G_{\leftrightarrow,\mathrm{dis}}(X,\prec)_S$ is connected; and (ii) for all $Y \in \Big(\mathrm{an}_G(S) \cap \mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)\Big) \setminus S$, there is no path from Y to any vertices in S. The condition (i) implies that $\mathrm{M}(X,\prec)$ will be large if the vertices in $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$ are connected by many bi-directed edges. The condition (ii) implies that $\mathrm{M}(X,\prec)$ will be large if there are few directed mixed cycles. Note that for ADMGs without directed mixed cycles, (ii) trivially holds since $\Big(\mathrm{an}_G(S) \cap \mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)\Big) \setminus S = \emptyset$. For example, consider a subset of vertices $\{V_1,\ldots,V_k\}$ in an ADMG with edges $V_i \leftrightarrow V_k, i = 1,\ldots,k-1$, which has no directed mixed cycles. Then, for an ordering $V_1 \prec \ldots \prec V_k$, $\mathrm{M}(V_k,\prec) = 2^{k-1}$. Also, consider a subset of vertices $\{V_1,\ldots,V_k\}$ in an ADMG with edges $V_1 \rightleftarrows V_2 \rightleftarrows \cdots \rightleftarrows V_k$, which has $k-1$ directed mixed cycles. Then, $\mathrm{M}(V_k,\prec) = 1$. Hence, it is clear that $\mathrm{M}(X,\prec)$ is large if

1. the set $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$ is large,

2. there are many bi-directed edges connecting vertices in $\mathrm{dis}_G(X) \cap \mathrm{pre}_{G,\prec}(X)$, and

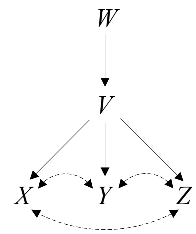


Figure 10: An example ADMG for which using (S-MP,$\prec$) is most beneficial. There is no directed mixed cycle and each c-component is a clique joined by bi-directed edges.

3. there are few directed mixed cycles.

Thus, (LMP,$\prec$) will invoke a large number of conditional independence relations for an ADMG with few directed mixed cycles and large c-components with many bi-directed edges. For such an ADMG, $\sum_{X\in \mathrm{C}(\prec)} \Big(\mathrm{M}(X,\prec)-1\Big)$, the reduction made by (S-MP,$\prec$), is also large. An extreme case is an ADMG that has no directed mixed cycles and each c-component of which is a clique joined by bi-directed edges. An example of such an ADMG is given in Figure 10. For this ADMG and an ordering $W \prec V \prec X \prec Y \prec Z$, (LMP,$\prec$) invokes $\mathrm{M}(W,\prec)+\mathrm{M}(V,\prec)+\mathrm{M}(X,\prec)+\mathrm{M}(Y,\prec)+\mathrm{M}(Z,\prec) = 1+1+1+2+4 = 9$ conditional independence relations while (S-MP,$\prec$) invokes $|\mathrm{C}(\prec)| = n = 5$ conditional independence relations. If we enlarge the clique joined by bi-directed edges such that it contains k vertices, then (LMP,$\prec$) invokes $2+\sum_{i=0}^{k-1} 2^i = 1+2^k$ conditional independence relations while (S-MP,$\prec$) invokes $k+2$.

In general, although (S-MP,$\prec$) greatly reduces (LMP,$\prec$), it may still invoke an exponential number of conditional independence relations if there exist directed mixed cycles.

5. Conclusion and Discussion

We present local Markov properties for ADMGs representing linear SEMs with correlated errors. The results have applications in testing linear SEMs against the data by testing for zero partial correlations implied by the model. For general linear SEMs with correlated errors, we provide a procedure that lists a subset of zero partial correlations that will imply all other zero partial correlations implied by the model. In particular, for a class of models whose corresponding path diagrams contain no directed mixed cycles, this subset invokes one zero partial correlation for each pair of variables.

In general, our procedure may invoke an exponential number of zero partial correlations if the path diagram G satisfies all of the following properties: (i) G has large c-components; (ii) the vertices in each c-component are heavily connected by bi-directed edges; and (iii) G has directed mixed cycles. If one of these properties is not satisfied, then the number of zero partial correlations derived by our method is typically not exponential.

For the class of MAGs, which is a strict superclass of ADMGs without directed mixed cycles, one might use the pairwise Markov property for MAGs given in Richardson and Spirtes (2002) instead of our results in Section 4. However, when the two approaches give a similar number of constraints, it may be better to use our approach since it may use smaller conditioning sets as shown in the example in Section 4.2.

The potential advantages of testing linear SEMs based on vanishing partial correlations over the classical test method based on maximum likelihood estimation of the covariance matrix have been discussed in Pearl (1998), Shipley (2000), McDonald (2002) and Shipley (2003). The results presented in this paper provide a theoretical foundation for the practical applications of this test method in linear SEMs with correlated errors. How to implement this test method in practice still needs further study as it requires multiple testing of hypotheses about zero partial correlations (Shipley, 2000; Drton and Perlman, 2007). We also note that, in linear SEMs *without* correlated errors, all the constraints on the covariance matrix are implied by vanishing partial correlations. This also holds in linear SEMs *with* correlated errors that are represented by ADMGs *without* directed mixed cycles. However, it is possible that linear SEMs *with* correlated errors represented by ADMGs *with* directed mixed cycles may imply constraints on the covariance matrix that are not implied by zero partial correlations.

Although the intended application is in linear SEMs, the local Markov properties presented in the paper are valid for ADMGs associated with any probability distributions that satisfy the composition axiom. For example, any probability distribution that is faithful[5] to some DAG or undirected graph (and the marginals of the distribution) satisfies the composition axiom.

Model debugging for ADMGs using vanishing partial correlations is another area of current research. In this model debugging problem, the goal is to modify a graph based on the pattern of rejected hypotheses. The properties of ADMGs presented in this paper may facilitate the development of a new model debugging method.

Acknowledgments

We thank the anonymous reviewers for helpful comments. This research was partly supported by NSF grant IIS-0347846.

References

S.A. Andersson, D. Madigan, and M.D. Perlman. Alternative Markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28:33–86, 2001.

K.A. Bollen. *Structural Equations with Latent Variables*. John Wiley, New York, 1989.

D.R. Cox and N. Wermuth. Linear dependencies represented by chain graphs. *Statistical Science*, 8(3):204–218, 1993.

M. Drton and M.D. Perlman. Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, 22(3):430–449, 2007.

O.D. Duncan. *Introduction to Structural Equation Models*. Academic Press, New York, 1975.

M. Frydenberg. The chain graph markov property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.

5. A probability distribution P is said to be faithful to a graph G if all the conditional independence relations embedded in P are encoded in G (via the global Markov property).

A.S. Goldberger. Structural equation models in the social sciences. *Econometrica: Journal of the Econometric Society*, 40:979–1001, 1972.

T. Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11: 1–12, 1943. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 477–490, 1995.

G. Kauermann. On a dualization of graphical Gaussian models. *Scandinavian Journal of Statistics*, 23:105–116, 1996.

J.T.A. Koster. On the validity of the Markov interpretation of path diagrams of gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, 26:413–431, 1999.

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

S.L. Lauritzen and N. Wermuth. Graphical models for association between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, 17:31–57, 1989.

S.L. Lauritzen, A.P. Dawid, B.N. Larsen, and H.G. Leimer. Independence properties of directed Markov fields. *Networks*, 20:491–505, 1990.

R.P. McDonald. What can we learn from the path equations?: Identifiability, constraints, equivalence. *Psychometrika*, 67(2):225–249, 2002.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, NY, 2000.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

J. Pearl. Graphs, causality, and structural equation models. *Socioligical Methods and Research*, 27: 226–284, 1998.

J. Pearl and P. Meshkat. Testing regression models with fewer regressors. In *Proceedings of AI-STAT*, pages 255–259. 1999.

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

T. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30(4):962–1030, 2002.

B. Shipley. A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, 7:206–218, 2000.

B. Shipley. Testing recursive path models with correlated errors using d-separation. *Structural Equation Modeling*, 10:214–221, 2003.

P. Spirtes, T. Richardson, C. Meek, R. Scheines, and C. Glymour. Using path diagrams as a structural equation modeling tool. *Socioligical Methods and Research*, 27:182–225, 1998.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2001.

N. Wermuth and D.R. Cox. Graphical models: overview. *International Encyclopedia of the Social and Behavioral Sciences*, 9:6379–6386, 2001.

N. Wermuth and D.R. Cox. Joint response graphs and separation induced by triangular systems. *Journal of the Royal Statistical Society B*, 66:687–717, 2004.

S. Wright. The method of path coefficients. *Ann. Math. Statist.*, 5:161–215, 1934.