Graphical Answers to Questions about Likelihood Inference for Gaussian Covariance Models*

Mathias Drton
University of California, Berkeley

Thomas S. Richardson
University of Washington

September 27, 2004

Abstract

In graphical modelling, a bi-directed graph encodes marginal independences among random variables that are identified with the vertices of the graph (alternatively graphs with dashed edges have been used for this purpose). Bi-directed graphs are special instances of ancestral graphs, which are mixed graphs with undirected, directed, and bi-directed edges. In this paper, we show how simplicial sets and the newly defined orientable edges can be used to construct a maximal ancestral graph that is Markov equivalent to a given bi-directed graph, i.e. the independence models associated with the two graphs coincide, and such that the number of arrowheads is minimal. Here the number of arrowheads of an ancestral graph is the number of directed edges plus twice the number of bi-directed edges. This construction yields an immediate check whether the original bi-directed graph is Markov equivalent to a directed acyclic graph (Bayesian network) or an undirected graph (Markov random field). Moreover, the ancestral graph construction allows for computationally more efficient maximum likelihood fitting of covariance graph models, i.e. Gaussian bi-directed graph models. In particular, we give a necessary and sufficient graphical criterion for determining when an entry of the maximum likelihood estimate of the covariance matrix must equal its empirical counterpart.

1 Introduction

In graphical modelling, a bi-directed graph encodes marginal independences among random variables that are identified with the vertices of the graph. In particular, whenever two vertices are not joined by an edge, then the two associated random variables are assumed to be marginally

*AMS 2000 subject classifications. Primary 62M45, 60K99; secondary 68R10, 68T30.
Keywords and Phrases: ancestral graphs, bi-directed graphs, covariance graphs, graphical models, marginal independence, maximum likelihood estimation.
Running head: Inference for Covariance Models

Figure 1: A bi-directed graph G encoding $X_1 \perp\!\!\!\perp X_2$, $X_1 \perp\!\!\!\perp X_4$, $X_2 \perp\!\!\!\perp X_3$, and its oriented simplicial graph G^{os}.

independent. In the Gaussian framework graphical models for marginal independence have been called covariance graph models (Cox and Wermuth, 1993, 1996) and are a special case of a linear hypothesis on the covariance matrix (Anderson, 1969, 1970, 1973). An estimation procedure especially designed for covariance graph models is described in Drton and Richardson (2003). The Markov interpretation of bi-directed graphs is discussed in Pearl and Wermuth (1994); Kauermann (1996); Banerjee and Richardson (2003); Richardson (2003). Examples of recent work involving graphical models for marginal independence are Grzebyk et al. (2004); Mao et al. (2004); Wermuth et al. (2004).

It should be noted that other authors (e.g. Cox and Wermuth, 1993, 1996; Edwards, 2000) have used dashed edges instead of bi-directed edges for the purpose of representing patterns of marginal independence. However, the bi-directed edges make explicit that graphical models for marginal independence are special instances of ancestral graphs (Richardson and Spirtes, 2002), which feature undirected, directed, and bi-directed edges. Through this a connection to path diagrams (e.g. Koster, 1999) and causality (Pearl, 2000; Spirtes et al., 2000; Richardson and Spirtes, 2003) is made.

As an example, let $X = (X_1, X_2, X_3, X_4)$ be a random vector in $\mathbb{R}^4$ that follows a centered multivariate normal distribution $\mathcal{N}(0, \Sigma)$ with positive definite covariance matrix $\Sigma \in \mathbb{R}^{4\times 4}$. Suppose that $X_1 \perp\!\!\!\perp X_2$, $X_1 \perp\!\!\!\perp X_4$, and $X_2 \perp\!\!\!\perp X_3$. Then Σ exhibits a pattern of zeroes as

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & \sigma_{13} & 0 \\ 0 & \sigma_{22} & 0 & \sigma_{24} \\ \sigma_{13} & 0 & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix}.$$

This zero pattern can be represented graphically by drawing a vertex for each one of the four random variables (labelled by the variable index) and joining vertex X_v and X_w by a bi-directed edge if $X_v \not\!\perp\!\!\!\perp X_w$, or equivalently under normality if σ_{vw} is not restricted to zero. In this example, we draw the graph G shown in Figure 1.

In this paper we employ the connection between bi-directed graphs for marginal independence and ancestral graphs (Section 2 introduces graphical concepts necessary for this connection). Using simplicial vertex sets (Section 3) and the newly defined orientable edges (Section 4), we show how to construct a maximal ancestral graph that is Markov equivalent to a given bi-directed graph, i.e. the independence models associated with the two graphs coincide, and such that the number of arrowheads is minimal. Here the number of arrowheads of an ancestral graph is the number of directed edges plus twice the number of bi-directed edges. We call the constructed ancestral graph an oriented simplicial graph (see e.g. Figure 1). This graph provides much useful information about the original bi-directed graph. It can be checked immediately whether the bi-directed graph is Markov equivalent to a directed acyclic graph, the type of graph that underlies Bayesian networks, or an undirected graph, the type of graph that underlies Markov random fields. For example, the graph G in Figure 1 is not Markov equivalent to an undirected graph because G^{os} is not an undirected graph, and G is not Markov equivalent to a directed acyclic graph because G^{os} contains a bi-directed edge (see also Pearl and Wermuth, 1994).

Employing information provided by oriented simplicial graphs in covariance graph models (Section 5) yields criteria for determining when an entry of the maximum likelihood estimate of the covariance matrix is available explicitly as its empirical counterpart. For example, the fact that no arrowheads appear at the vertices 1 and 2 in the graph G^{os} in Figure 1 implies that the maximum likelihood estimates of σ_{11} and σ_{22} must be equal to the empirical variance of X_1 and X_2, respectively. In addition, oriented simplicial graphs allow for computationally more efficient maximum likelihood fitting (compare Drton and Richardson, 2003, §4.2.4). We conjecture that the ancestral graph construction will also provide useful information in future work on bi-directed graph models for discrete random variables.

2 Graphical terminology

2.1 Mixed graphs

Commonly a *graph* $G = (V, E)$ consists of a finite *vertex set* V and an *edge set* $E \subseteq V \times V$. If $(v, w) \in E$ and $(w, v) \in E$, then the pictorial representation of G contains an undirected edge $v - w$. If $(v, w) \in E$ but $(w, v) \notin E$, then the edge is directed as $v \to w$. This definition of graphs is the traditional basis for graphical models (Lauritzen, 1996, Chapter 2). In this paper, however, we will consider graphs that feature undirected and directed edges in conjunction with bi-directed edges, drawn as $v \leftrightarrow w$. Graphs with these three types of edges are called mixed graphs and can be defined formally as follows (cf. Richardson and Spirtes, 2002, Appendix).

Let $\mathcal{E} = \{-, \leftarrow, \to, \leftrightarrow\}$ be the set of possible edges between an ordered pair of vertices. A *mixed graph* $G = (V, E)$ is a pair of a finite *vertex set* V and an *edge mapping* $E : V \times V \to \mathcal{P}(\mathcal{E})$, where $\mathcal{P}(\mathcal{E})$ denotes the power set of $\mathcal{E}$. The edge mapping E has to satisfy that for all $v, w \in V$,

(i) $E(v, v) = \emptyset$, i.e. there is no edge between a vertex and itself,

(ii) $- \in E(v, w) \iff - \in E(w, v)$,

(iii) $\rightarrow \in E(v,w) \iff \leftarrow \in E(w,v)$,

(iv) $\leftrightarrow \in E(v,w) \iff \leftrightarrow \in E(w,v)$,

In this paper, we consider exclusively *simple mixed graphs*, for which there is at most one edge between two vertices, i.e. $|E(v,w)| \leq 1$ for all $v, w \in V$.

For a mixed graph $G = (V, E)$ and vertices $v, w \in V$, we write $v - w \in G$, $v \rightarrow w \in G$, $v \leftarrow w \in G$ or $v \leftrightarrow w \in G$ if $-$, $\rightarrow$, $\leftarrow$ or $\leftrightarrow$ are in $E(v,w)$, respectively. If $E(v,w) \neq \emptyset$, then v and w are *adjacent*. If there is an edge $v \leftarrow w \in G$ or $v \leftrightarrow w \in G$ then there is said to be an *arrowhead at* v on this edge. If there is an edge $v \rightarrow w \in G$ or $v - w \in G$ then there is said to be a *tail at* v on this edge. A vertex w is said to be in the *boundary* of v if v and w are adjacent. We denote the boundary of v as $\mathrm{bd}(v)$. The boundary of vertex set $A \subseteq V$ is the set

$$\mathrm{bd}(A) = \cup(\mathrm{bd}(v) \mid v \in A) \setminus A.$$

We write $\mathrm{Bd}(v) = \mathrm{bd}(v) \cup \{v\}$ and $\mathrm{Bd}(A) = \mathrm{bd}(A) \cup A$. An induced subgraph of G over a vertex set A is the mixed graph $G_A = (A, E_A)$ where E_A is the restriction of the edge mapping E on $A \times A$.

In a simple mixed graph a sequence of adjacent vertices $(v_1, \ldots, v_k)$ uniquely determines the sequence of edges joining consecutive vertices v_i and v_{i+1}, $1 \leq i \leq k-1$. Hence, we can define a *path* π between two vertices v and w in a simple mixed graph as a sequence of distinct vertices $\pi = (v, v_1, \ldots, v_k, w)$ such that each vertex in the sequence is adjacent to its predecessor and its successor. A path of the form $v \rightarrow \cdots \rightarrow w$, on which every edge is of the form $\rightarrow$, with the arrowheads pointing toward w, is a directed path from v to w. If there is such a directed path from v to $w \neq v$, or if $v = w$, then v is an *ancestor* of w. We denote the set of all ancestors of a vertex v by $\mathrm{an}(v)$ and for a vertex set $A \subseteq V$ we define

$$\mathrm{an}(A) = \cup(\mathrm{an}(v) \mid v \in A).$$

Finally, a directed path from v to w together with an edge $w \rightarrow v \in G$ is called a *directed cycle*.

Important subclasses of simple mixed graphs are the following. *Bi-directed graphs* feature only bi-directed edges $\leftrightarrow$, and *undirected graphs* only undirected edges $-$. *Directed acyclic graphs* (DAGs), also called *acyclic directed graphs*, have only directed edges $\rightarrow$ or $\leftarrow$ such that there are no directed cycles. These three types of graphs are contained in the class of *ancestral graphs* (Richardson and Spirtes, 2002), where G is an ancestral graph if the following conditions are met:

(i) if $v - w \in G$, then there does not exist u such that $u \rightarrow v \in G$ or $u \leftrightarrow v \in G$;

(ii) there are no directed cycles;

(iii) if $v \leftrightarrow w \in G$, then v is not an ancestor of w.

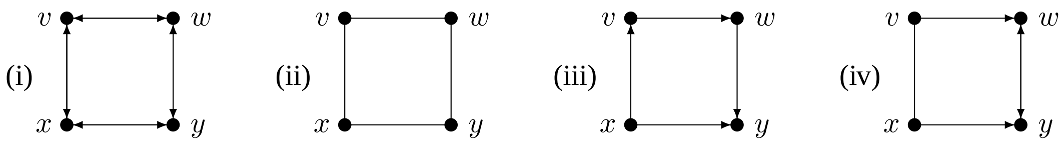

Figure 2: Examples of simple mixed graphs. (i) A bi-directed graph, (ii) an undirected graph, (iii) a DAG, (iv) an ancestral graph that does not fall into one of the previous three classes of graphs.

Figure 2 shows examples of the above graph types. As mentioned in the introduction, bi-directed graphs encode patterns of marginal independence. Undirected graphs encode the conditional independence patterns underlying Markov random fields. DAGs are the graphs that underlie Bayesian networks. The motivation for the introduction of ancestral graphs was to find a class of graphs that contains all DAGs while having a Markov property (see next section) that is closed under conditioning and marginalization. In particular, for every ancestral graph there exists a DAG such that the ancestral graph encodes the pattern of conditional independence arising from the DAG in a conditioning and marginalization process. This way ancestral graphs inherit causal interpretability from DAGs (Richardson and Spirtes, 2003).

2.2 Global Markov property for ancestral graphs and Markov equivalence

The global Markov property for ancestral graphs is based on the m-separation criterion (Richardson and Spirtes, 2002, §3.4), which naturally extends Pearl's (1988) d-separation criterion for DAGs to ancestral graphs.

A nonendpoint vertex v on a path is a *collider on the path* if the edges preceding and succeeding v on the path both have an arrowhead at v, that is, $\rightarrow v \leftarrow$, $\rightarrow v \leftrightarrow$, $\leftrightarrow v \leftarrow$ or $\leftrightarrow v \leftrightarrow$ is part of the path. A nonendpoint vertex v on a path which is not a collider is a *non-collider on the path*. A path between vertices v and w in an ancestral graph G is said to be *m-connecting given a set $C \subseteq V$* (possibly empty), with $v, w \notin C$, if:

(i) every non-collider on the path is not in C, and

(ii) every collider on the path is in $\mathrm{an}(C)$.

If no path m-connects v and w given C, then v and w are said to be m-separated given C. Sets A and B are m-separated given C, if for every pair v, w, with $v \in A$ and $w \in B$, v and w are m-separated given C (A, B, C are disjoint sets; A, B are nonempty).

Now let the vertex set of an ancestral graph $G = (V, E)$ index a family of random variables $(X_v \mid v \in V)$. For $A \subseteq V$, let X_A be the random vector $(X_v \mid v \in A)$. For disjoint sets $A, B, C \subseteq V$, $A, B \neq \emptyset$, the *global Markov property* for the ancestral graph G states that $X_A \perp\!\!\!\perp X_B \mid X_C$, i.e. X_A is conditionally independent of X_B given X_C, whenever A and B are

m-separated given C in G. Subsequently, we will write $A \perp\!\!\!\perp B \mid C$ as a shorthand that avoids making the probabilistic context explicit. Further details on Markov properties of graphs can be found in particular in Lauritzen (1996); Cowell et al. (1999); Edwards (2000).

For the graphs shown in Figure 2, the global Markov property of

(i) the bi-directed graph states $v \perp\!\!\!\perp y$ and $w \perp\!\!\!\perp x$;

(ii) the undirected graph states $v \perp\!\!\!\perp y \mid (w, x)$ and $w \perp\!\!\!\perp x \mid (v, y)$;

(iii) the DAG states $v \perp\!\!\!\perp y \mid (w, x)$ and $w \perp\!\!\!\perp x \mid v$;

(iv) the ancestral graph states $v \perp\!\!\!\perp y \mid x$ and $w \perp\!\!\!\perp x \mid v$.

It is clear that bi-directed graphs encode marginal independences in that their global Markov property states the pairwise marginal independence $v \perp\!\!\!\perp w$ if v and w are not adjacent. It can be shown that in a multivariate normal distribution these pairwise marginal independences hold iff all independences stated by the global Markov property of the bi-directed graph hold (see Kauermann, 1996; Banerjee and Richardson, 2003). Without making a particular distributional assumption, Richardson (2003, §4) shows that the independences stated by the global Markov property of a bi-directed graph hold iff certain (not only pairwise) marginal independences hold.

The global Markov property of the ancestral graph in Figure 2(iv) associates an independence statement with every edge that is missing in the graph. This, however, is not true for all ancestral graphs (see Richardson and Spirtes, 2002, §3.7). If an ancestral graph G is such that the global Markov property states that v and w are conditionally independent given some C whenever v and w are not adjacent, then G is a *maximal ancestral graph*. Fortunately, for every non-maximal ancestral graph G there exists a unique Markov equivalent maximal ancestral graph $\bar{G}$ such that G is a subgraph of $\bar{G}$. Here two ancestral graphs G_1 and G_2 are *Markov equivalent* if the global Markov property for G_1 states the same independences as the global Markov property for G_2. Note that bi-directed and undirected graphs as well as DAGs are always maximal.

Subsequently we will employ repeatedly the following simple Lemma.

Lemma 2.1. *If G is a bi-directed graph and $\bar{G}$ is an ancestral graph such that G and $\bar{G}$ have the same skeleton and are Markov equivalent, then $\bar{G}$ is a maximal ancestral graph.*

Proof. If v and w are two non-adjacent vertices in $\bar{G}$, then they are also non-adjacent in G, which implies that $v \perp\!\!\!\perp w$ is stated by the global Markov property of G because in G every non-endpoint vertex on a path is a collider. Since $\bar{G}$ is Markov equivalent to G it follows that $v \perp\!\!\!\perp w$ is also implied by the global Markov property of $\bar{G}$. Thus, $\bar{G}$ is maximal. □

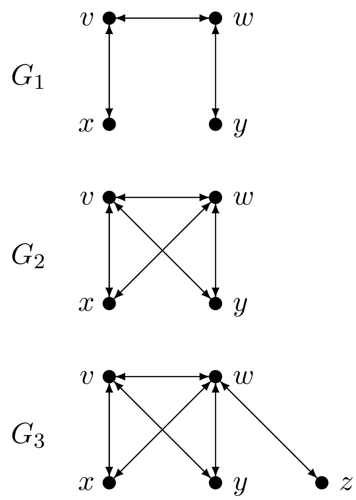

Figure 3: Examples of bi-directed graphs and their simplicial graphs.

3 Simplicial sets and Markov equivalence

In this section we show how simplicial vertex sets of a bi-directed graph can be used to construct a Markov equivalent maximal ancestral graph by removing arrowheads from certain bi-directed edges.

Definition 3.1 (Simplicial vertices and sets). *A vertex $v \in V$ is* simplicial*, if* $\mathrm{Bd}(v)$ *is complete. Similarly, a set $A \subseteq V$ is* simplicial*, if* $\mathrm{Bd}(A)$ *is complete.*

Simplicial sets are also important in other contexts of graphical modelling; for example in collapsibility (Madigan and Mosurski, 1990; Kauermann, 1996; Lauritzen, 1996, §2.1.3, p.121 and 219) and triangulation of graphs (Jensen, 2001, §5.3).

If there is an arrowhead at v on an edge between v and w, then we say that we *drop the arrowhead at* v when either $v \to w$ is replaced by $v - w$ or $v \leftrightarrow w$ is replaced by $v \to w$.

Definition 3.2 (Simplicial graph). *Let G be a bi-directed graph and G^s the mixed graph obtained by dropping all the arrowheads at simplicial vertices of G. We call G^s the* simplicial graph *induced by the bi-directed graph G.*

Figure 3 shows the bi-directed graphs G_1, G_2, and G_3 with their simplicial graphs G_1^s, G_2^s, and G_3^s, respectively. As shown in the Theorem 3.3, a simplicial graph is a maximal ancestral graph that is Markov equivalent to the bi-directed graph that induced it.

Theorem 3.3 (Markov equivalence of the simplicial graph). *Let G^s be the* simplicial graph *induced by a bi-directed graph G. Then* *(i)* G^s *is a maximal ancestral graph;*

(ii) G *and* G^s *are Markov equivalent;*

(iii) G *is Markov equivalent to an undirected graph iff* G^s *is an undirected graph iff* G *is a disjoint union of complete graphs.*

Proof. (i) Let $v \in V$. First assume that there exists $w \in V$ such that $v - w \in G^s$. By definition of G^s, v must be simplicial. Hence, there may not exist an arrowhead at v, which means that there may not exist $u \in V$ such that either $v \leftarrow u \in G^s$ or $v \leftrightarrow u \in G^s$.

Next assume that either $u \to v \in G^s$ or $u \leftrightarrow v \in G^s$ and that there exists a directed path $v \to v_1 \to \cdots \to v_k \to u \in G^s$. However, the presence of the directed edge $v \to v_1 \in G^s$ implies that v is simplicial, which is in contradiction to the fact that there is an arrowhead at v on the edge between v and u. Thus, G^s is an ancestral graph.

Finally, G^s is a maximal ancestral graph by Lemma 2.1 in conjunction with the Markov equivalence established in (ii) below.

(ii) Since two vertices are adjacent in G^s iff they are adjacent in G, the claim follows if we can show that two non-adjacent vertices v and w are m-connected given $C \subseteq V$ in G iff they are m-connected given C in G^s.

First, let v and w be two non-adjacent vertices that are m-connected given $C \subseteq V$ in G. Let $\pi = (v, v_1, \ldots, v_k, w)$ be a path in G that is of minimal length among all paths in G that m-connect v and w given C. This implies that $v_1, \ldots, v_k$ are not simplicial. Moreover, because G is a bi-directed graph, $v_1, \ldots, v_k$ are colliders and $\{v_1 \ldots, v_k\} \subseteq C$; compare Richardson (2003, Lemma 6(iii) and (iv)). Since two vertices are adjacent in G^s iff they are adjacent in G, there exists a (unique) path $\pi^s = (v, v_1, \ldots, v_k, w)$ in G^s. Since $v_1, \ldots, v_k$ are not simplicial in G, they are colliders on π^s, and $\{v_1 \ldots, v_k\} \subseteq C$ yields that π^s is m-connecting v and w in G^s.

Conversely, let v and w be two vertices that are m-connected given $C \subseteq V$ in G^s and let $\pi^s = (v = v_0, v_1, \ldots, v_k, v_{k+1} = w)$ be a path in G^s that is of minimal length among all paths in G^s that m-connect v and w given C. Assume there exists a simplicial vertex v_i on π^s. Then it follows that v_{i-1} and v_{i+1} are adjacent in G^s, and that $\pi^s_{-i} = (v, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k, w)$ is a path in G^s. By definition of G^s, a vertex u in G^s is either such that every edge with end-point u has an arrowhead at u, or such that every edge with end-point u has a tail at u. This implies that π^s_{-i} m-connects v and w given C contradicting that π^s is the shortest such path. Therefore, all $v_1, \ldots, v_k$ are non-simplicial vertices and thus colliders on π^s. Moreover, no vertex among $v_1, \ldots, v_k$ is ancestral to a vertex in C. This implies that $\{v_1 \ldots, v_k\} \subseteq C$, which yields that the path $\pi = (v, v_1, \ldots, v_k, w)$ in G is m-connecting v and w given C.

(iii) If G^s is an undirected graph, then by (ii) G is Markov equivalent to an undirected graph, namely G^s. Conversely if G^s is not an undirected graph, assume that there exists an undirected graph U that is Markov equivalent to G. Necessarily, G, G^s and U have the same skeleton. Since G^s is not undirected, there exists a vertex v that is not simplicial, i.e. there exist two non-adjacent vertices u and w in $\mathrm{bd}(v)$. The global Markov property for G states that $u \perp\!\!\!\perp w$.

However, the path (u, v, w) m-connects u and w given the empty set in U. Thus, the global Markov property of U does not state $u \perp\!\!\!\perp w$, which contradicts the assumption that U and G are Markov equivalent.

Finally the simplicial graph G^s is an undirected graph iff the vertex set of the inducing bi-directed graph G can be partitioned into pairwise disjoint sets $A_1, \ldots, A_q$ such that (a) if $v \in A_i$, $1 \leq i \leq q$, and $w \in A_j$, $1 \leq j \leq q$, are adjacent, then $i = j$, and (b) all the induced subgraphs G_{A_i}, $i = 1, \ldots, q$ are complete graphs (Kauermann, 1996). □

Remark 3.4. The simplicial graph G^s sometimes may be a DAG. For example, let G be the bi-directed graph over the vertex set $V = \{u, v, w\}$ that has two edges as $u \leftrightarrow v \leftrightarrow w$. The Markov property of this graph states $u \perp\!\!\!\perp w$ and the simplicial graph G^s is equal to the DAG $u \rightarrow v \leftarrow w$. However, there exist bi-directed graphs that are Markov equivalent to a DAG but for which the simplicial graph still contains bi-directed edges. For example, the graph G_2 in Figure 3 is equivalent to the DAG obtained from G_2^s by replacing the bi-directed edge $v \leftrightarrow w$ by either $v \rightarrow w$ or $v \leftarrow w$. We pursue this further in the next section.

4 Orientable edges and Markov equivalence

4.1 Orientable edges

In this section, we develop a general notion of orientable edges that will allow us to orient some of the bi-directed edges in a simplicial graph, that is replace them by directed edges, such that the new graph with fewer bi-directed edges is still a maximal ancestral graph that is Markov equivalent to the original bi-directed graph. In fact, the resulting graph will minimize the number of arrowheads in the class of maximal ancestral graphs that are Markov equivalent to the original bi-directed graph.

Definition 4.1 (Orientable edges). *An edge between two vertices v and w is*

(i) orientable as $v \rightarrow w$ *if* $\mathrm{Bd}(v) \subseteq \mathrm{Bd}(w)$*;*

(ii) uniquely orientable as $v \rightarrow w$ *if* $\mathrm{Bd}(v) \subsetneq \mathrm{Bd}(w)$*;*

(iii) covered *if it is both orientable as $v \rightarrow w$ and orientable as $v \leftarrow w$, or equivalently if* $\mathrm{Bd}(v) = \mathrm{Bd}(w)$*;*

(iv) non-orientable *if it is neither orientable as $v \rightarrow w$ nor orientable as $v \leftarrow w$.*

Lemma 4.2 (Picture for non-orientable edges). *Let G be a bi-directed graph. An edge between v and w in a bi-directed graph G is* not *orientable iff there exist vertices $x \in \mathrm{bd}(v) \setminus \{w\}$ and $y \in \mathrm{bd}(w) \setminus \{v\}$ such that the induced subgraph $G_{\{x,y,v,w\}}$ equals one of the two graphs shown in Figure 4.*

Figure 4: Induced subgraphs of a bi-directed graph with non-orientable edge between v and w.

Proof. The claim follows immediately from Definition 4.1; compare also Pearl and Wermuth (1994). $\square$

Lemma 4.3 (Transitivity of orientability). *Let u, v and w be three distinct vertices. If there are edges between u and v, and v and w that are orientable as $u \to v$ and $v \to w$, respectively, then there is also an edge between u and w, which is orientable as $u \to w$.*

Proof. By definition 4.1(i), $\mathrm{Bd}(u) \subseteq \mathrm{Bd}(v) \subseteq \mathrm{Bd}(w)$, which implies the presence of an edge between u and w because $u \in \mathrm{Bd}(v) \subseteq \mathrm{Bd}(w)$. This edge is orientable as $u \to w$ because $\mathrm{Bd}(u) \subseteq \mathrm{Bd}(w)$. $\square$

The binary relation between vertices based on orientability is not antisymmetric in general since covered edges may exist; e.g. the edge $v \leftrightarrow w$ in the graph G_2 in Figure 3 is covered. Thus, it is not a partial order. However, the binary relation based on unique orientability does provide a partial order on the vertex set of a graph.

Definition 4.4 (Unique orientability relation). *For two vertices v and w, we write $v \preceq w$ if either $v = w$ or v and w are joined by an edge that is uniquely orientable as $v \to w$. We call $\preceq$ the* unique orientability relation*.*

Lemma 4.5 (Partial order). *The unique orientability relation is a partial order on the vertex set V.*

Proof. (reflexivity): By definition, $v \preceq v$ for all $v \in V$.

(antisymmetry): If $v \preceq w$ and $w \preceq v$ but $v \neq w$, then $\mathrm{Bd}(v) \subsetneq \mathrm{Bd}(w)$ and $\mathrm{Bd}(w) \subsetneq \mathrm{Bd}(v)$, which is a contradiction.

(transitivity): Let $u \preceq v$ and $v \preceq w$. Then clearly, $u \preceq w$ if $u = v$ or $v = w$. Otherwise u, v, and w are three distinct vertices and there are edges between u and v, and between v and w that are uniquely orientable as $u \to v$ and $v \to w$, respectively. Thus $\mathrm{Bd}(u) \subsetneq \mathrm{Bd}(v) \subsetneq \mathrm{Bd}(w)$. Since $u \in \mathrm{Bd}(v)$, it follows that $u \in \mathrm{Bd}(w)$, i.e. there is an edge between u and w, which must be uniquely orientable as $u \to w$ because $\mathrm{Bd}(u) \subsetneq \mathrm{Bd}(w)$. $\square$

4.2 Orientable edges in the simplicial graph

An edge between two vertices v and w in a bi-directed graph G is orientable as $v \to w$ iff it is orientable as $v \to w$ in the simplicial graph G^s that is induced by G. However, more can be said about the orientable edges in G^s.

Lemma 4.6 (Edges in the simplicial graph). *Let G be a bi-directed graph with simplicial graph G^s and let v and w be adjacent. Then*

(i) if $v - w \in G^s$, then this edge is covered;

(ii) if $v \to w \in G^s$, then this edge is uniquely orientable as $v \to w$;

(iii) if $v \leftrightarrow w \in G^s$, then this edge may be covered, uniquely orientable as $v \to w$ or $v \leftarrow w$, or non-orientable.

Proof. (i) If $v - w \in G^s$, then both v and w are simplicial. Since $w \in \mathrm{Bd}(v)$, and $\mathrm{Bd}(v)$ is complete, it follows that $\mathrm{Bd}(v) \subseteq \mathrm{Bd}(w)$. Similarly, $\mathrm{Bd}(v) \supseteq \mathrm{Bd}(w)$.

(ii) If $v \to w \in G^s$, then v is simplicial and $\mathrm{Bd}(v) \subseteq \mathrm{Bd}(w)$ follows from the argument in the proof of (i). Thus, the edge $v \to w \in G^s$ is orientable as $v \to w$. Next, assume that the edge $v \to w \in G^s$ is also orientable as $v \leftarrow w$. This means that $\mathrm{Bd}(w) \subseteq \mathrm{Bd}(v)$. Since there is an arrowhead at w on the edge $v \to w \in G^s$, w is not simplicial. Therefore, there exist $x, y \in \mathrm{bd}(w)$ such that x and y are not adjacent in G^s. But $x, y \in \mathrm{bd}(w) \subseteq \mathrm{Bd}(v)$, hence $\mathrm{Bd}(v)$ is not complete contradicting $v \to w \in G^s$. Therefore, the edge $v \to w \in G^s$ is not orientable as $v \leftarrow w$.

(iii) In the graph G_1^s in Figure 3, the edge $v \leftrightarrow w$ is non-orientable; compare Lemma 4.2. In the graph G_2^s in Figure 3 $v \leftrightarrow w$ is covered, whereas in the graph G_3^s in Figure 3 it is uniquely orientable as $v \to w$. □

4.3 Oriented simplicial graphs

We now show how a maximal ancestral graph can be constructed by replacing orientable bi-directed edges in the simplicial graph by directed edges while preserving Markov equivalence to the original bi-directed graph.

Definition 4.7 (Oriented simplicial graph). *Let G^s be the simplicial graph of a bi-directed graph G. Create a new graph G^{os} from G^s as follows:*

(i) replace every bi-directed edge in G^s that is uniquely orientable as $v \to w$ by the directed edge $v \to w$;

(ii) replace every covered bi-directed edge in G^s by a directed edge such that G^{os} contains no directed cycles.

We call G^{os} an oriented simplicial graph *induced by the bi-directed graph G.*

Such an oriented simplicial graph will always exist: it can be constructed using the fact that the unique orientability relation is a partial order (Lemma 4.5). This fact implies that the vertex set of G can be *well-ordered* as $V = \{v_1, \ldots, v_p\}$ such that

$$v_i \preceq v_j \Longrightarrow i \leq j.$$

It follows from Lemma 4.6(ii) that after the introduction of the directed edges in step (i) of Definition 4.7, an edge $v_i \rightarrow v_j$ can only occur if $i < j$. Furthermore, in step (ii) we can select the directed edges such that $v_i \rightarrow v_j \in G^{os}$ only if $i < j$. Then G^{os} does not contain any directed cycles and meets the criterion to be an oriented simplicial graph.

Example G_2 in Figure 3 shows that there may indeed exist multiple oriented simplicial graphs induced by a given bi-directed graph because replacing the edge $v \leftrightarrow w$ in G_2^s by either $v \rightarrow w$ or $v \leftarrow w$ will result in an oriented simplicial graph.

Lemma 4.8 (Edges in the oriented simplicial graph). *Let G be a bi-directed graph and G^{os} be an induced oriented simplicial graph. It holds that*

(i) if $v - w$ is an undirected edge in G^{os}, then it is covered;

(ii) if $v \rightarrow w$ is a directed edge in G^{os}, then it is orientable as $v \rightarrow w$;

(iii) $v \leftrightarrow w$ is a bi-directed edge in G^{os} iff it is non-orientable.

Proof. (i) An edge in G^{os} is undirected iff this edge is already undirected in G^s, and the claim follows from Lemma 4.6(i).

(ii) An edge in G^{os} is directed either if it is already directed in G^s, or if it is bi-directed and orientable in G^s. Thus the claim follows from Lemma 4.6(ii).

(iii) The claim follows immediately from Definition 4.7. □

Figure 5 shows oriented simplicial graphs for the three bi-directed graphs G_1, G_2, and G_3 whose simplicial graphs we depicted in Figure 3.

Theorem 4.9 (Markov equivalence of an oriented simplicial graph). *Let G be a bi-directed graph and G^{os} be an induced oriented simplicial graph. Then*

(i) G^{os} is a maximal ancestral graph;

(ii) G and G^{os} are Markov equivalent;

(iii) G is Markov equivalent to an undirected graph iff G^{os} is an undirected graph iff G is a disjoint union of complete graphs;

(iv) G is Markov equivalent to a DAG iff G^{os} contains no bi-directed edges;

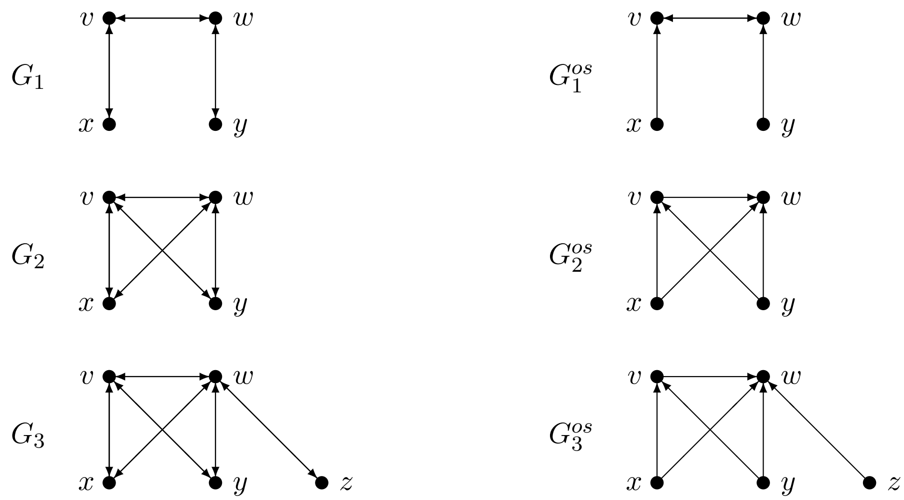

Figure 5: Examples of oriented simplicial graphs.

(v) *G^{os} has the minimum number of arrowheads of all maximal ancestral graphs that are Markov equivalent to G. Here the number of arrowheads of an ancestral graph G with d directed and b bi-directed edges is defined as* $\mathrm{arr}(G) = d + 2b$.

Proof of Theorem 4.9(i). Since $v - w \in G^{os}$ iff $v - w \in G^s$ it follows that there may not exist an arrowhead at v or w; compare the proof of Theorem 3.3(i). Furthermore, by definition, G^{os} does not contain any directed cycles. Finally, assume that there exists $v \leftrightarrow w \in G^{os}$ and a directed path $\pi = (v = v_0, v_1, \ldots, v_k, v_{k+1} = w)$ from v to w. Then by Lemma 4.8, all the edges $v_i \to v_{i+1}$ on π are orientable as $v_i \to v_{i+1}$. By transitivity of orientability (Lemma 4.3), it follows that there is an edge $v \to v_k \in G^{os}$ that is orientable as $v \to v_k$. The transitivity then implies that the edge $v \leftrightarrow w \in G^{os}$ is orientable as $v \to w$, which contradicts the assumption that $v \leftrightarrow w \in G^{os}$, which occurs iff the edge between v and w is non-orientable; compare Lemma 4.8(iii). In conclusion, we have shown that G^{os} is an ancestral graph. The maximality of G^{os} will follow from the proof of Theorem 4.9(ii) and Lemma 2.1. □

In the proofs of parts (ii)-(v) in Theorem 4.9, we will make use of results on m-connecting paths in an oriented simplicial graph that are obtained in Appendix A.

Proof of Theorem 4.9(ii). We proceed by elaborating the proof of Theorem 3.3(ii).

First, let v and w be two non-adjacent vertices that are m-connected given $C \subseteq V$ in G. Let $\pi = (v = v_0, v_1, \ldots, v_k, v_{k+1} = w)$ m-connect v and w given C in G and be such that no shorter path m-connects v and w given C. Then $v_1, \ldots, v_k$ are colliders, $\{v_1 \ldots, v_k\} \subseteq C$, and v_{i-1} and v_{i+1}, $i = 1, \ldots, k$, are not adjacent in G. Therefore, for all $i = 1, \ldots, k-1$, $v_{i-1} \in \mathrm{Bd}(v_i)$ but $v_{i-1} \notin \mathrm{Bd}(v_{i+1})$, and similarly $v_{i+2} \notin \mathrm{Bd}(v_i)$ but $v_{i+2} \in \mathrm{Bd}(v_{i+1})$. It follows that $\mathrm{Bd}(v_i) \nsubseteq \mathrm{Bd}(v_{i+1})$ and $\mathrm{Bd}(v_i) \nsupseteq \mathrm{Bd}(v_{i+1})$ for all $i = 1, \ldots, k-1$, which yields that all the edges $v_i \leftrightarrow v_{i+1} \in G$, $i = 1, \ldots, k-1$, are non-orientable. Thus, by Lemma 4.8(iii), $v_i \leftrightarrow v_{i+1} \in G^{os}$, $i = 1, \ldots, k-1$, and all $v_2, \ldots, v_{k-1}$ are colliders on the path $\pi^{os} = (v, v_1, \ldots, v_k, w)$ in G^{os}. In addition it follows similarly that $v_2 \in \mathrm{Bd}(v_1) \setminus \mathrm{Bd}(v)$, which entails that $\mathrm{Bd}(v_1) \nsubseteq \mathrm{Bd}(v)$ and that the edge between v and v_1 is not orientable as $v \leftarrow v_1$. Thus, v_1 is a collider on π^{os}. Analogously, we can show that v_k is a collider on π^{os}, which concludes the proof of the fact that π^{os} is a path in G^{os} that m-connects v and w given C.

Conversely, let v and w be two vertices that are m-connected given $C \subseteq V$ in G^{os}. By Proposition A.5 from the Appendix, there exists a path $\pi^{os} = (v, v_1, \ldots, v_k, w)$ that m-connects v and w given C in G^{os} and is such that $v_1, \ldots, v_k$ are colliders with $\{v_1 \ldots, v_k\} \subseteq C$. This entails that the path $\pi = (v, v_1, \ldots, v_k, w)$ in G m-connects v and w given C. □

Proof of Theorem 4.9(iii). The claim follows immediately from Theorem 3.3(iii) since G^{os} is an undirected graph iff G^s is an undirected graph. □

Proof of Theorem 4.9(iv). Let G be a bi-directed graph such that G^{os} does not contain any bi-directed edges. It follows from Definition 3.1 that the induced subgraph G^{os}_A is undirected and complete if $A \subseteq V$ is a simplicial set. Let $A_1, \ldots, A_q$ be the inclusion-maximal simplicial sets of G. Let D be a directed graph obtained by replacing each induced subgraph $G^{os}_{A_i}$, $i = 1, \ldots, q$, by a complete DAG. Then D itself has to be acyclic, i.e. a DAG, which can be seen as follows: First, since G^{os} is an ancestral graph and thus does not contain any directed cycles, a directed cycle π in D must involve a vertex $v \in \cup(A_i \mid i = 1, \ldots, q)$. Let $v \in A_j$. Since the induced subgraphs D_{A_i}, $i = 1, \ldots, q$, are all acyclic, π must also involve a vertex not in A_j. Therefore, there must exist an edge $x \rightarrow w$ on π such that $w \in A_j$ and $x \notin A_j$. Since the sets A_i are inclusion-maximal simplicial sets, no vertex in A_i, $i \neq j$, is adjacent to any vertex in A_j. Hence, $x \notin \cup(A_i \mid i = 1, \ldots, q)$, which implies that the edge $x \rightarrow w$ is also present in G^{os}. The presence of an arrowhead at w in G^{os}, however, implies that w is not a simplicial vertex, and thus w cannot be in A_j, which is a contradiction.

Now two vertices are adjacent in G^{os} iff they are adjacent in D. If two vertices v and w are not adjacent and $\pi = (v, v_1, \ldots, v_k, w)$ is a shortest m-connecting path given C in G^{os}, then, by Proposition A.5 from the Appendix, $\pi = (v, v_1, w)$ with v_1 being a collider; otherwise there would have to be a bi-directed edge in G^{os}. In particular, $v_1 \notin \cup(A_i \mid i = 1, \ldots, q)$ and thus, (v, v_1, w) has identical edge structure in D and G^{os}. Similarly, a path in D that is shortest among paths m-connecting two vertices v and w given C is of the form $\pi = (v, v_1, w)$ with $v_1 \notin \cup(A_i \mid i = 1, \ldots, q)$ being a collider. Thus, (v, v_1, w) has identical edge structure in G^{os} and in particular is also m-connecting in G^{os}, which concludes the proof of Markov equivalence of G^{os} and the DAG D.

Conversely, let G be a bi-directed graph with an oriented simplicial graph G^{os} such that $v \leftrightarrow w \in G^{os}$. Suppose for a contradiction that G is equivalent to a DAG D. Note that D must have the same skeleton as G and thus as G^{os}. By Lemma 4.8(iii), $v \leftrightarrow w$ is non-orientable.

Hence, by Lemma 4.2, there exist $x \in \mathrm{bd}(v) \setminus \{w\}$ and $y \in \mathrm{bd}(w) \setminus \{v\}$. By the global Markov property of G, $x \perp\!\!\!\perp w$ and $v \perp\!\!\!\perp y$. It follows that on the path (x, v, w) in D, v must be a collider. Therefore, $v \leftarrow w \in D$. But then w is a non-collider on the path (v, w, y) in D contradicting the independence $v \perp\!\!\!\perp y$. □

Remark 4.10. The proof of Theorem 4.9(iv) in particular provides a proof of the Markov equivalence result stated without proof in Pearl and Wermuth (1994, Thm. 1) and Drton and Richardson (2003, Prop. 2).

Proof of Theorem 4.9(v). Let $\bar{G}$ be a maximal ancestral graph that is Markov equivalent to the (bi-directed) graph G. The graphs $\bar{G}$, G, and G^{os} must have the same skeleton. Assume that $\mathrm{arr}(\bar{G}) < \mathrm{arr}(G^{os})$. Then either (a) there exists $v \to w \in G^{os}$ such that $v - w \in \bar{G}$ or (b) there exists $v \leftrightarrow w \in G^{os}$ such that $v \to w \in \bar{G}$ or $v - w \in \bar{G}$.

Case (a): If $v \to w \in G^{os}$, then w cannot be simplicial. Hence, there exist $x, y \in \mathrm{bd}(w)$ such that x and y are not adjacent in G^{os}, thus not adjacent in G. The global Markov property of G states that $x \perp\!\!\!\perp y$. Since $\bar{G}$ is an ancestral graph and $v - w \in \bar{G}$, however, there may not be any arrowheads at w on the edges between x and w, and y and w in $\bar{G}$. Therefore, x and y are m-connected given $\emptyset$ in $\bar{G}$, which yields that the global Markov property of $\bar{G}$ does not imply $x \perp\!\!\!\perp y$; a contradiction.

Case (b): Now $v \leftrightarrow w \in G^{os}$ but there is no arrowhead at v on the edge between v and w in $\bar{G}$. By Lemma 4.8(iii) $v \leftrightarrow w$ is non-orientable in G^{os}, and by Lemma 4.2 there exists $x \in \mathrm{bd}(v) \setminus \mathrm{Bd}(w)$ such that x and w are not adjacent in G^{os}. Thus x and w are not adjacent in G and $x \perp\!\!\!\perp w$ is stated by the global Markov property of G. In $\bar{G}$, however, v is a non-collider on the path (x, v, w) and thus this path m-connects x and w given $\emptyset$, which yields that the global Markov property of $\bar{G}$ does not imply $x \perp\!\!\!\perp w$; a contradiction. □

5 Maximum likelihood estimation in covariance graph models

5.1 Covariance graph models

Let G be a bi-directed graph, and

$$\mathbf{P}(G) = \left\{ \Sigma \in \mathbb{R}^{V \times V} \mid \Sigma = (\sigma_{vw}) \text{ sym. pos. def.}, \ \sigma_{vw} = 0 \ \forall (v, w) : v \leftrightarrow w \notin G \right\} \tag{5.1}$$

be the cone of symmetric positive definite matrices with a zero pattern induced by G. Then

$$\mathbf{N}(G) = \left(\mathcal{N}_V(0, \Sigma) \mid \Sigma \in \mathbf{P}(G) \right) \tag{5.2}$$

is the *covariance graph model* associated with G. It can be shown that every distribution in $\mathbf{N}(G)$ already satisfies all conditional independences stated by the global Markov property for the bi-directed graph G (Kauermann, 1996, Prop. 2.2).

Assume that $S \in \mathbb{R}^{V\times V}$ is the empirical covariance matrix computed from an i.i.d. sample drawn from some unknown distribution $\mathcal{N}_V(0,\Sigma) \in \mathbf{N}(G)$, i.e. the (v,w)-th entry in S is the dot product of the vectors of observations for the v-th and w-th variables divided by the sample size n. For empirical covariance matrix S and sample size n, the log-likelihood function $\ell_{S,n} : \mathbf{P}(G) \to \mathbb{R}$ of $\mathbf{N}(G)$ can be written as

$$\ell_{S,n}(\Sigma) = -\frac{nV}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{n}{2}\mathrm{tr}\left(\Sigma^{-1}S\right). \tag{5.3}$$

Here V also denotes the cardinality of the vertex set V. If S is positive definite then the global maximum of $\ell_{S,n}$ over $\mathbf{P}(G)$ exists and thus Σ can be estimated by the maximizer of $\ell_{S,n}$. The likelihood equations obtained by setting to zero the partial derivatives of $\ell_{S,n}$ with respect to the non-restricted entries in Σ are non-linear and take on the form

$$(\Sigma^{-1})_{vw} = (\Sigma^{-1}S\Sigma^{-1})_{vw} \quad \forall v,w \in V: \ v = w \ \text{ or } \ v \leftrightarrow w \in G; \tag{5.4}$$

compare Drton and Richardson (2003, §3.2). A positive definite matrix $\hat{\Sigma}(S) \in \mathbf{P}(G)$ that solves the equations (5.4) is called *a solution to the likelihood equations* of $\mathbf{N}(G)$. By a connection between conditional likelihood functions in covariance graph models and the likelihood functions of seemingly unrelated regressions (Drton and Richardson, 2003, §4.2), it follows from Drton (2004); Drton and Richardson (2004b) that for a given positive definite empirical covariance matrix S there may exist multiple solutions to the likelihood equations (5.4) and the likelihood functions may exhibit multiple local maxima.

5.2 Parameterization of Gaussian ancestral graph models

Since subsequent theorems are obtained via results in Gaussian ancestral graph models, we briefly review the parameterization of these models.

Let G be an ancestral graph. The parameterization associates one parameter with each vertex in V and each edge in E, respectively. Due to property (i) in the definition of an ancestral graph (see Section 2.1), we can define $\mathrm{un}_G \subseteq V$ as the set of vertices v that are such that any edge with endpoint v has a tail at v. Clearly $v - w \in G$ implies $v, w \in \mathrm{un}_G$, and $v \leftrightarrow w \in G$ implies that $v, w \notin \mathrm{un}_G$. Let Λ be a symmetric positive definite $\mathrm{un}_G \times \mathrm{un}_G$ matrix such that $\Lambda_{vw} \neq 0$ only if $v = w$ or $v - w \in G$. Let Ω be a symmetric positive definite $(V \setminus \mathrm{un}_G) \times (V \setminus \mathrm{un}_G)$ matrix such that $\Omega_{vw} \neq 0$ only if $v = w$ or $v \leftrightarrow w \in G$. Finally, let B be a $V \times V$ matrix such that $B_{vw} \neq 0$ only if $w \to v \in G$. With the parameter matrices Λ, B, Ω, we can define the symmetric positive definite matrix

$$\Sigma(\Lambda, B, \Omega) = (I_V - B)^{-1}\begin{pmatrix}\Lambda^{-1} & 0\\ 0 & \Omega\end{pmatrix}(I_V - B)^{-\mathrm{T}}. \tag{5.5}$$

Let $\mathbf{N}(G)$ be the Gaussian ancestral graph model associated with G, i.e. the family of all centered normal distributions that are globally Markov with respect to G. As shown in Richardson and Spirtes (2002, §8), the normal distribution $\mathcal{N}_V(0,\Sigma)$ with $\Sigma = \Sigma(\Lambda, B, \Omega)$ defined in (5.5) is in $\mathbf{N}(G)$. Conversely, if G is a *maximal* ancestral graph, then for any $\mathcal{N}_V(0,\Sigma) \in \mathbf{N}(G)$ there exist unique Λ, Ω, B of the above type such that $\Sigma = \Sigma(\Lambda, B, \Omega)$. Note that if G is a bi-directed graph then the definition of $\mathbf{N}(G)$ made here agrees with the definition in (5.2). Further details on the parameterization can be found in Richardson and Spirtes (2002, §8); Drton and Richardson (2004a). Note, however, that Richardson and Spirtes (2002) use B for what is here denoted $I_V - B$.

Now let G be a bi-directed graph and G^{os} an induced oriented simplicial graph. Since G and G^{os} are Markov equivalent the reparameterization mapping $(\Lambda, B, \Omega) \mapsto \Sigma(\Lambda, B, \Omega)$ has image equal to $\mathbf{P}(G)$. By Richardson and Spirtes (2002, Thm. 8.14, Lemma 8.22), the mapping is a diffeomorphism, which implies the following Lemma.

Lemma 5.1. *If G is a bi-directed graph with oriented simplicial graph G^{os}, then (Λ, B, Ω) solves the likelihood equations of $\mathbf{N}(G^{os})$ iff $\Sigma(\Lambda, B, \Omega)$ solves the likelihood equations of $\mathbf{N}(G)$.*

5.3 Empirical maximum likelihood estimates of covariances

Employing the graphical results established in Sections 3 and 4, we show next that for a simplicial set A in a bi-directed graph G, the $A \times A$ submatrix of a solution to the likelihood equations (5.4) must agree with the counterpart in the empirical covariance matrix S.

Theorem 5.2 (Empirical covariances). *Let G be a bi-directed graph with associated covariance graph model $\mathbf{N}(G)$. If $A \subseteq V$ is simplicial, S is a symmetric positive definite matrix, and $\hat{\Sigma}(S) \in \mathbf{P}(G)$ is a solution to the likelihood equations (5.4), then*

$$\hat{\Sigma}(S)_{A\times A} = S_{A\times A}. \tag{5.6}$$

Proof. Let G^s be the simplicial graph induced by G. By Theorem 4.9(ii), the covariance graph model $\mathbf{N}(G)$ and the Gaussian ancestral graph model $\mathbf{N}(G^s)$ based on G^s contain exactly the same multivariate normal distributions. Let $\mathbf{N}(G^s)$ be parameterized by the precision matrix Λ, the matrix of regression coefficients B and the covariance matrix Ω as described in Section 5.2. In particular, it follows from Richardson and Spirtes (2002, Lemma 8.4) that if $\Sigma = \Sigma(\Lambda, B, \Omega)$, then

$$(\Lambda^{-1})_{A\times A} = \Sigma_{A\times A}. \tag{5.7}$$

Let $A_1, \dots, A_q$ be the inclusion-maximal simplicial sets of G. Then these sets are pairwise disjoint and the induced subgraphs $G^s_{A_i}$, $i = 1, \dots, q$, complete undirected graphs. In fact, $\cup_i A_i = \mathrm{un}_{G^s}$. It follows that Λ is a block-diagonal matrix such that $\Lambda_{vw} = 0$ if there does not exist an inclusion-maximal simplicial set A_i such that $v \in A_i$ and $w \in A_i$. Now the discussion in Richardson and Spirtes (2002, §8.5) and Lemma 5.1 imply that every solution to the likelihood equations for Λ, B, Ω in the Gaussian ancestral graph model $\mathbf{N}(G^s)$ satisfies

$$(\Lambda^{-1})_{A_i\times A_i} = S_{A_i\times A_i} \quad \forall i = 1,\dots,q,$$

and, in particular, since $A \subseteq A_j$ for some j,

$$(\Lambda^{-1})_{A\times A} = S_{A\times A} \quad \forall i = 1,\dots,q.$$

In conjunction with Lemma 5.1 and (5.7), this yields the claim. □

Theorem 5.2 implies in particular that for a simplicial vertex $v \in V$ a global maximum $\hat{\Sigma}(S) \in \mathbf{P}(G)$ of the likelihood function of $\mathbf{N}(G)$ must satisfy $\hat{\Sigma}(S)_{vv} = S_{vv}$. In general this is false for non-simplicial vertices as the next theorem shows.

Theorem 5.3 (Non-empirical variances). *If $v \in V$ is* not *simplicial, then there exists a symmetric positive definite matrix S such that any global maxima $\hat{\Sigma}(S) \in \mathbf{P}(G)$ of the likelihood function of $\mathbf{N}(G)$ satisfies*

$$\hat{\Sigma}(S)_{vv} \neq S_{vv}.$$

Proof. Since v is not simplicial, there exist non-adjacent vertices u and w such that $u \leftrightarrow v \in G$ and $v \leftrightarrow w \in G$. Let $R = V \setminus \{u, w, v\}$, and choose S as the block-diagonal matrix

$$S = \begin{array}{c} \\ R \\ u \\ w \\ v \end{array} \begin{array}{c} \begin{array}{cccc} R & u & w & v \end{array} \\ \begin{pmatrix} I_R & 0_R & 0_R & 0_R \\ 0_R' & 3 & 1 & 1 \\ 0_R' & 1 & 3 & 1 \\ 0_R' & 1 & 1 & 3 \end{pmatrix} \end{array}, \tag{5.8}$$

where I_R is the $R\times R$ identity matrix, and 0_R is the zero vector in $\mathbb{R}^R$, i.e. $0_R' = (0,\dots,0)$. Clearly, S is symmetric and positive definite.

In the following we will relax the optimization problem of maximizing the likelihood function $\ell_{S,n}$ over $\mathbf{P}(G)$ by working with a more simply structured supergraph of G. Let $\bar{G}$ be the bi-directed graph that has all possible edges except for the edge $u \leftrightarrow w$; see Figure 6. Let $\hat{\Sigma}(S, \bar{G})$ be a global maximum of $\ell_{S,n}(\Sigma)$ over $\mathbf{P}(\bar{G})$ and let $\hat{\Sigma}(S, G) = \hat{\Sigma}(S)$ be a global maximum of $\ell_{S,n}(\Sigma)$ over $\mathbf{P}(G)$. Clearly, $\mathbf{P}(G) \subseteq \mathbf{P}(\bar{G})$ and thus

$$\ell_{S,n}(\hat{\Sigma}(S,G)) = \max_{\Sigma\in\mathbf{P}(G)} \ell_{S,n}(\Sigma) \leq \max_{\Sigma\in\mathbf{P}(\bar{G})} \ell_{S,n}(\Sigma) = \ell_{S,n}(\hat{\Sigma}(S,\bar{G})). \tag{5.9}$$

In the bi-directed graph $\bar{G}$, $\mathrm{Bd}(u) = V \setminus \{w\}$, $\mathrm{Bd}(w) = V \setminus \{u\}$ and $\mathrm{Bd}(x) = V$ if $x \notin \{u, w\}$. Thus, an edge between x and y in $\bar{G}$ is covered if $x, y \notin \{u, w\}$, and it is uniquely orientable as $x \to y$ if $x \in \{u, w\}$ and $y \notin \{u, w\}$. Thus an oriented simplicial graph induced

Figure 6: The supergraph $\bar{G}$ and its oriented simplicial graph $\bar{G}^{os}$.

by $\bar{G}$ cannot have any bi-directed edges (Lemma 4.8), which implies that $\bar{G}$ is equivalent to a DAG (Theorem 4.9). This DAG can be chosen as a DAG D such that every edge with endpoint u and w has a tail at u and w, respectively, and every edge with endpoint v has an arrowhead at v; see Figure 6.

Standard results on maximum likelihood estimation in Gaussian DAG models (e.g. Andersson and Perlman, 1998, compare also Theorem 5.4 below) yield that

$$\hat{\Sigma}(S, \bar{G}) = \begin{array}{c} \\ R \\ u \\ w \\ v \end{array} \begin{array}{c} \begin{array}{cccc} R & u & w & v \end{array} \\ \begin{pmatrix} I_R & 0_R & 0_R & 0_R \\ 0_R' & 3 & 0 & 0.75 \\ 0_R' & 0 & 3 & 0.75 \\ 0_R' & 0.75 & 0.75 & 2.875 \end{pmatrix} \end{array} \tag{5.10}$$

is the unique local and also global maximum of $\ell_{S,n}$ over $\mathbf{P}(\bar{G})$. Since $\hat{\Sigma}(S, \bar{G}) \in \mathbf{P}(G)$ it follows from (5.9) that $\hat{\Sigma}(S, G) = \hat{\Sigma}(S, \bar{G})$ is the unique global maximum of $\ell_{S,n}$ over $\mathbf{P}(G)$, which in conjunction with (5.10) proves the claim. $\square$

5.4 Empirical maximum likelihood estimates of conditional parameters

Oriented simplicial graphs also provide information on when maximum likelihood estimates of conditional parameters are equal to empirical counterparts. The conditional parameters we consider are regression coefficients and conditional variance for the conditional distribution of variable v given its *parents*

$$\mathrm{pa}(v) = \mathrm{pa}_{G^{os}}(v) = \{w \in V \mid w \to v \in G^{os}\}.$$

If $\mathrm{pa}(v) = \emptyset$, then conditioning variable v on the empty set is understood to yield the marginal distribution of v.

Theorem 5.4 (Empirical conditional parameters). *Let G^{os} be an oriented simplicial graph induced by the bi-directed graph G, S a symmetric positive definite matrix, and $\hat{\Sigma}(S) \in \mathbf{P}(G)$ a solution to the likelihood equations (5.4). Moreover, let v be such that no vertex w is adjacent to v by a bi-directed edge $v \leftrightarrow w \in G^{os}$. It then holds that the regression coefficients equal*

$$\hat{\Sigma}(S)_{v \times \mathrm{pa}(v)} \left[\hat{\Sigma}(S)_{\mathrm{pa}(v) \times \mathrm{pa}(v)}\right]^{-1} = S_{v \times \mathrm{pa}(v)} \left(S_{\mathrm{pa}(v) \times \mathrm{pa}(v)}\right)^{-1}, \tag{5.11}$$

and that the conditional variance equals

$$\hat{\Sigma}(S)_{vv} - \hat{\Sigma}(S)_{v\times \mathrm{pa}(v)}\big[\hat{\Sigma}(S)_{\mathrm{pa}(v)\times \mathrm{pa}(v)}\big]^{-1}\hat{\Sigma}(S)_{\mathrm{pa}(v)\times v} = \\ S_{vv} - S_{v\times \mathrm{pa}(v)}\big(S_{\mathrm{pa}(v)\times \mathrm{pa}(v)}\big)^{-1}S_{\mathrm{pa}(v)\times v}. \quad (5.12)$$

Proof. If $\mathrm{pa}(v) = \emptyset$, then v is a simplicial vertex, and the claim reduces to $\hat{\Sigma}(S)_{vv} = S_{vv}$, which follows from Theorem 5.2. Otherwise, using the parameterization of $\mathbf{N}(G^{os})$, it follows from Richardson and Spirtes (2002, Thm. 8.7) that if $\Sigma = \Sigma(\Lambda, B, \Omega)$, then

$$\Sigma_{v\times \mathrm{pa}(v)}\big[\Sigma_{\mathrm{pa}(v)\times \mathrm{pa}(v)}\big]^{-1} = B_{v\times \mathrm{pa}(v)}$$

and

$$\Sigma_{vv} - \Sigma_{v\times \mathrm{pa}(v)}\big[\Sigma_{\mathrm{pa}(v)\times \mathrm{pa}(v)}\big]^{-1}\Sigma_{\mathrm{pa}(v)\times v} = \Omega_{vv}.$$

Now if Λ, B, Ω solve the likelihood equations for $\mathbf{N}(G^{os})$, then $B_{v\times \mathrm{pa}(v)}$ and Ω_{vv} solve the likelihood equations of the model in which all parameters in Λ, B, Ω except for $B_{v\times \mathrm{pa}(v)}$ and Ω_{vv} are treated as fixed values. It now follows from Drton and Richardson (2004a, §§5.1 and 5.3) that $B_{v\times \mathrm{pa}(v)}$ and Ω_{vv} must be unique for all solutions to the likelihood equations of $\mathbf{N}(G^{os})$ and equal to the empirical expressions on the right hand side of (5.11) and (5.12), respectively. Applying Lemma 5.1 thus yields the claim. □

Remark 5.5. If $\mathrm{pa}(v) = \emptyset$, then v is simplicial iff it satisfies the assumptions of Theorem 5.4. Therefore, Theorem 5.3 implies that the statement of Theorem 5.4 is false in general if v is the endpoint of a bi-directed edge.

Remark 5.6 (Efficient iterative conditional fitting on the oriented simplicial graph). The algorithm for maximum likelihood estimation proposed in Drton and Richardson (2003) is a special purpose algorithm for maximum likelihood estimation of the covariance matrix in a covariance graph model, but the algorithm does not make use of the information provided by Theorems 5.2 and 5.4. An alternative strategy is to run the extension of the algorithm to Gaussian ancestral graph models described in Drton and Richardson (2004a) on an oriented simplicial graph. This extended algorithm then avoids unnecessary computations as it will implicitly make use of Theorems 5.2 and 5.4.

5.5 Unsolvability by radicals of the likelihood equations

Theorem 5.3 shows that if v is the endpoint of a bi-directed edge then, in a covariance graph model, the maximum likelihood estimator of the variance Σ_{vv} may not be equal to the empirical variance. However, as Theorem 5.7 below shows, it may also be the case that the solutions to the likelihood equations yielding an estimate of Σ_{vv} cannot be computed from the data in finitely many steps involving addition, subtraction, multiplication, division, or taking n-th roots. Hence, the conditions given in Theorems 5.2 and 5.4 under which closed form estimates exist are both necessary and sufficient.

Let G be the bi-directed graph in Figure 4(i). If one conditions v and w on x and y in the covariance graph model $\mathbf{N}(G)$, then one obtains a set of conditional distributions that form a bivariate seemingly unrelated regressions model. For this model, Drton and Richardson (2004b) show that solving the likelihood equations is equivalent to computing the roots of a quintic polynomial. Furthermore, they give example data for which this quintic has exactly three real roots. Galois theory (Stewart, 1989, Lemma 14.7) now implies that for these data the quintic is unsolvable by radicals, i.e. the roots of the quintic and thus the solutions to the likelihood equations cannot be computed from the data in finitely many steps involving addition, subtraction, multiplication, division, or taking n-th roots. (See Geiger et al. (2002) for similar results in the context of undirected graphs.)

Theorem 5.7 (Unsolvability by radicals). *Let G^{os} be an oriented simplicial graph induced by the bi-directed graph G. Moreover, let v be such that there exists a bi-directed edge $v \leftrightarrow w \in G^{os}$. Then there exists a symmetric positive definite matrix S for which the global maximum $\hat{\Sigma}(S) \in \mathbf{P}(G)$ of the likelihood function of $\mathbf{N}(G)$ cannot be computed in finitely many steps involving addition, subtraction, multiplication, division, or taking n-th roots.*

Proof. If $v \leftrightarrow w \in G^{os}$, then by Lemma 4.2 there exist two vertices x and y such that the induced subgraph $G_{\{x,y,v,w\}}$ is one of the graphs in Figure 4. Let $R = V \setminus \{x, y, v, w\}$ and choose S such that $S_{uu} = 1$ if $u \in R$, $S_{uz} = 0$ if $u, z \in R$ but $u \neq z$. The submatrix $S_{R\times R}$ choose equal to the sample covariance matrix of the data in Drton and Richardson (2004b, Table 1). Considering the bi-directed graph $\bar{G}$, in which only the edges between v and y, and x and w are absent, instead of the graph G yields analogously to the proof of Theorem 5.3 that the $\{x, y, v, w\} \times \{x, y, v, w\}$ submatrix of $\hat{\Sigma}(S)$ can be computed by fitting the model based on the graph in Figure 4(i) to the $\{x, y, v, w\} \times \{x, y, v, w\}$ submatrix of S. Hence, the discussion right before Theorem 5.7 applies and yields the claim. □

Remark 5.8 (Dual estimation). Even though the likelihood equations of covariance graph models may sometimes be unsolvable in the sense of Theorem 5.7, the dual estimator of Kauermann (1996) may still be available as rational function of the data. This occurs in the case where the skeleton of the bi-directed graph, i.e. the undirected graph with the same adjacencies, is decomposable. For example, the graph in Figure 4(i) has a decomposable skeleton.

A Connecting paths in oriented simplicial graphs

In this appendix we prove results used in the proof of Theorem 4.9.

Let v and w be some fixed vertices that are m-connected given $C \subseteq V$ in an oriented simplicial graph G^{os}. Let $\Pi^{os}(v, w|C)$ be the set of paths that m-connect v and w given C in

G^{os}, and let $\Pi^{os}_{\min}(v, w|C)$ be the set of paths that are of minimal length among the paths in $\Pi^{os}(v, w|C)$.

Lemma A.1. *If v_{i-1}, v_i and v_{i+1} are three consecutive vertices on a path π in G^{os}, and v_i is a non-collider on π, then v_{i-1} and v_{i+1} are adjacent.*

Proof. If v_i is a non-collider, then the edge between v_i and v_{i-1} or the edge between v_i and v_{i+1} must have a tail at v_i. Without loss of generality, assume that there is a tail on the edge between v_i and v_{i+1}. Then by Lemma 4.6(i) and (ii), this edge is orientable as $v_i \rightarrow v_{i+1}$. Hence, $\mathrm{Bd}(v_i) \subseteq \mathrm{Bd}(v_{i+1})$, from which it follows that $v_{i-1} \in \mathrm{Bd}(v_{i+1})$, which is the claim. □

Lemma A.2. *If $\pi = (v = v_0, v_1, \ldots, v_k, v_{k+1} = w) \in \Pi^{os}(v, w|C)$, and v_i is a non-endpoint vertex on π and there is an arrowhead at v_i on the edge between v_{i-1} and v_i, then either (i) $v_i \in \mathrm{an}(C)$ or (ii) the path $(v_i, v_{i+1}, \ldots, v_k, w)$ is a directed path from v_i to w.*

Proof. Suppose the result is false. Let v_j be the vertex closest to w satisfying the antecedent of the Lemma, but not the conclusion. If v_j is a collider, then by definition of m-connection, $v_j \in \mathrm{an}(C)$, which is a contradiction. If v_j is a non-collider then $v_j \rightarrow v_{j+1}$ on π. If $v_{j+1} = w$, if $v_{j+1} \in \mathrm{an}(C)$, or if $(v_{j+1}, \ldots, v_k, w)$ is a directed path from v_{j+1} to w, then clearly v_j satisfies the conclusion of the Lemma, which is a contradiction. But if $v_{j+1} \notin \mathrm{an}(C) \cup \{w\}$ and $(v_{j+1}, \ldots, v_k, w)$ is not a directed path from v_{j+1} to w then v_{j+1} satisfies the conditions on v_j, but is closer to w, again a contradiction. □

Lemma A.3. *If $\pi = (v = v_0, v_1, \ldots, v_k, v_{k+1} = w) \in \Pi^{os}_{\min}(v, w|C)$ then no non-consecutive vertices on π are adjacent.*

Proof. Suppose for a contradiction that there are non-consecutive vertices on the path which are adjacent. Let (v_p, v_q) be a pair of vertices which are furthest apart on the path, i.e.

$$(p, q) \in \big\{(r, s) \in \{0, \ldots, k+1\}^2 \mid r < s,\ |r - s| > 1,\ v_r \text{ and } v_s \text{ are adjacent in } G^{os}\big\}.$$

If $(v_p, v_q) = (v, w)$ then clearly π is not minimal. Hence either $v \neq v_p$ or $w \neq v_q$.

Suppose that $v \neq v_p$. By definition of (p, q), v_{p-1} is not adjacent to v_q. Consequently, by Lemma A.1, v_q is a collider on (v_{p-1}, v_p, v_q), and thus the edge between v_{p-1} and v_p has an arrowhead at v_p. It then follows by Lemma A.2 that either $v_p \in \mathrm{an}(C)$ or $(v_p, v_{p+1}, \ldots, v_k, w)$ is a directed path from v_p to w. In the latter case $v_p \in \mathrm{an}(v_q)$, but there is an arrowhead at v_p on the edge between v_p and v_q, which contradicts that G^{os} is ancestral. Hence $v_p \in \mathrm{an}(C)$. If $v_q = w$ then the path $(v, v_1, \ldots, v_p, v_q = w)$ is m-connecting given C and shorter than π. Hence $v_q \neq w$. It then follows by the same argument that v_q is a collider on (v_p, v_q, v_{q+1}) and in $\mathrm{an}(C)$. However, this also leads to a contradiction since then the path $(v, v_1, \ldots, v_p, v_q, v_{q+1}, \ldots, v_k, w)$ is both m-connecting given C and shorter than π.

The case where $w \neq v_q$ may be argued symmetrically. □

Corollary A.4. *If $\pi = (v = v_0, v_1, \dots, v_k, v_{k+1} = w) \in \Pi^{os}_{\min}(v, w|C)$, then all the non-endpoint vertices $v_1, \dots, v_k$ are colliders on π.*

Proof. This follows directly from Lemma A.3 and Lemma A.1. □

Even though all non-endpoints on a path in $\Pi^{os}_{\min}(v, w|C)$ are colliders, not all non-endpoints must be in the set C. For example, in the graph G^{os}_2 from Figure 5, the path (x, v, y) m-connects x and y given w since the collider v is ancestral to w. However, as the next Proposition shows, there will always exist a path in $\Pi^{os}_{\min}(v, w|C)$ such that all non-endpoints are colliders in C. In G^{os}_2 from Figure 5, the path (x, w, y) m-connects x and y given $\{w\}$.

Proposition A.5. *Let $\pi = (v = v_0, v_1, \dots, v_k, v_{k+1} = w) \in \Pi^{os}_{\min}(v, w|C)$ be such that no other path in $\Pi^{os}_{\min}(v, w|C)$ has more non-endpoint vertices in C than π. Then all non-endpoint vertices $v_1, \dots, v_k$ on π are colliders that are in C.*

Proof. By Corollary A.4, all non-endpoints $v_1, \dots, v_k$ are colliders. Assume that there exists $v_i \notin C$, $1 \leq i \leq k$. Since $\pi \in \Pi^{os}(v, w|C)$, and thus $v_i \in \mathrm{an}(C)$, there exists $c \in C$ such that $v_i \to \dots \to c \in G^{os}$. In particular, $c \neq v_{i-1}$ and $c \neq v_{i+1}$ because v_i is ancestral neither to v_{i-1} nor to v_{i+1}. Lemma 4.8(ii), transitivity of orientability (Lemma 4.3), and the fact that G^{os} does not contain directed cycles imply that $v_i \to c \in G^{os}$. By Lemma A.1, G^{os} contains edges between c and both v_{i-1} and v_{i+1}. Since the edge between v_{i-1} and v_i has an arrowhead at v_i and $v_i \to c \in G^{os}$, the edge between v_{i-1} and c must have an arrowhead at c because otherwise the fact that G^{os} is an ancestral graph would be contradicted (Theorem 4.9(i)). Similarly, the edge between v_{i+1} and c must have an arrowhead at c. If $v_{i-1} \to c$, then v_{i-2} is adjacent to c and by the same argument as above there must be an arrowhead at c on the edge between v_{i-2} and c. Repeating this argument yields that there exists a vertex v_ℓ, $\ell \leq i - 1$, such that either $v_\ell \leftrightarrow c \in G^{os}$, or $v_\ell = v$ and $v \to c$. The same arguments also imply that there exists a vertex v_j, $j \geq i + 1$, such that either $v_j \leftrightarrow c \in G^{os}$, or $v_j = w$ and $w \to c$. Therefore, the path $(v, v_1, \dots, v_\ell, c, v_j, \dots, v_k, w)$ is in $\Pi^{os}(v, w|C)$ and either shorter than π or of equal length but with more non-endpoint vertices in C. This contradicts the choice of π and therefore the assumption of a non-endpoint on π that is not in C must be false. □

References

ANDERSON, T. W. (1969). Statistical inference for covariance matrices with linear structure. In *Multivariate Analysis, II (Proc. Second Internat. Sympos., Dayton, Ohio, 1968)*. Academic Press, New York, 55–66.

ANDERSON, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In *Essays in Probability and Statistics*. University of North Carolina Press, Chapel Hill, N.C., 1–24.

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

ANDERSSON, S. A. and PERLMAN, M. D. (1998). Normal linear regression models with recursive graphical Markov structure. *J. Multivariate Anal.* **66** 133–187.

BANERJEE, M. and RICHARDSON, T. S. (2003). On a dualization of graphical Gaussian models: A correction note. *Scand. J. Statist.* **30** 817–820.

COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1999). *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science, Springer-Verlag, New York.

COX, D. R. and WERMUTH, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statist. Sci.* **8** 204–218,247–277.

COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. Chapman and Hall, London.

DRTON, M. (2004). Computing all roots of the likelihood equations of seemingly unrelated regressions. *Journal of Symbolic Computation* Submitted.

DRTON, M. and RICHARDSON, T. S. (2003). A new algorithm for maximum likelihood estimation in Gaussian graphical models for marginal independence. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence* (U. Kjærulff and C. Meek, eds.). Morgan Kaufmann, San Francisco, 184–191.

DRTON, M. and RICHARDSON, T. S. (2004a). Iterative conditional fitting for Gaussian ancestral graph models. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (M. Chickering and J. Halpern, eds.). Morgan Kaufmann, San Francisco, 130–137.

DRTON, M. and RICHARDSON, T. S. (2004b). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. *Biometrika* **91** 383–392.

EDWARDS, D. M. (2000). *Introduction to Graphical Modelling*, 2nd ed. Springer-Verlag, New York.

GEIGER, D., MEEK, C. and STURMFELS, B. (2002). On the toric algebra of graphical models. Tech. Rep. MSR-TR-2002-47, Microsoft Research. Submitted to the *Annals of Statistics*.

GRZEBYK, M., WILD, P. and CHOUANIÈRE, D. (2004). On identification of multi-factor models with correlated residuals. *Biometrika* **91** 141–151.

JENSEN, F. V. (2001). *Bayesian Networks and Decision Graphs*. Springer-Verlag, New York.

KAUERMANN, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Statist.* **23** 105–116.

KOSTER, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equation systems with correlated errors. *Scand. J. Statist.* **26** 413–431.

LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series, Clarendon Press, Oxford, UK.

MADIGAN, D. and MOSURSKI, K. (1990). An extension of the results of Asmussen and Edwards on collapsibility in contingency tables. *Biometrika* **77** 315–319.

MAO, Y., KSCHISCHANG, F. R. and FREY, B. J. (2004). Convolutional factor graphs as probabilistic models. In *Procceding of the 20th Conference on Uncertainty in Artificial Intelligence* (U. Kjærulff and C. Meek, eds.). Morgan Kaufmann, San Francisco, 374–381.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

PEARL, J. (2000). *Causality*. Cambridge University Press, Cambridge, UK.

PEARL, J. and WERMUTH, N. (1994). When can association graphs admit a causal interpretation? In *Selecting Models from Data: Artificial Intelligence and Statistics IV*, vol. 89 of *Lecture Notes in Statistics*. Springer, New York, 205–214.

RICHARDSON, T. S. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Statist.* **30** 145–157.

RICHARDSON, T. S. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030.

RICHARDSON, T. S. and SPIRTES, P. (2003). Causal inference via ancestral graph models. In *Highly Structured Stochastic Systems* (P. Green, N. Hjort and S. Richardson, eds.), chap. 3. Oxford University Press, Oxford, UK, 83–105.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. MIT Press, Cambridge, MA.

STEWART, I. (1989). *Galois Theory*, 2nd ed. Chapman and Hall, London.

WERMUTH, N., COX, D. R. and MARCHETTI, G. (2004). Decompositions and estimation of a chain of covariances. Manuscript.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
1073 EVANS HALL #3840
BERKELEY, CA, 94720-3840
U.S.A.
E-MAIL: drton@math.berkeley.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WA 98105-4322
U.S.A.
E-MAIL: tsr@stat.washington.edu