*J. R. Statist. Soc.* B (2004)
**66**, *Part* 3, *pp.* 687–717

# Joint response graphs and separation induced by triangular systems

Nanny Wermuth

*Gothenburg University, Sweden*

and D. R. Cox

*Nuffield College, Oxford, UK*

[Received October 2001. Final revision December 2003]

**Summary.** We consider joint probability distributions generated recursively in terms of univariate conditional distributions satisfying conditional independence restrictions. The independences are captured by missing edges in a directed graph. A matrix form of such a graph, called the generating edge matrix, is triangular so the distributions that are generated over such graphs are called triangular systems. We study consequences of triangular systems after grouping or reordering of the variables for analyses as chain graph models, i.e. for alternative recursive factorizations of the given density using joint conditional distributions. For this we introduce families of linear triangular equations which do not require assumptions of distributional form. The strength of the associations that are implied by such linear families for chain graph models is derived. The edge matrices of chain graphs that are implied by any triangular system are obtained by appropriately transforming the generating edge matrix. It is shown how induced independences and dependences can be studied by graphs, by edge matrix calculations and via the properties of densities. Some ways of using the results are illustrated.

*Keywords*: Chain graphs; Concentration graphs; Covariance graphs; Directed acyclic graphs; Graphical Markov models; Univariate recursive regressions

## 1. Introduction

### 1.1. *Triangular systems and induced chain graph models*

Analysis and interpretation of multivariate data can often be simplified when knowledge about independences is available or can be derived. A summary of such independences can be given in graphs in which variables are represented by nodes. Conditional associations between pairs of variables can then be represented by edges between nodes and independences by missing edges. For different perspectives on the statistical models that are associated with such independence graphs, see Cox and Wermuth (1996), Edwards (2000), Green *et al.* (2003), Lauritzen (1996) and Whittaker (1990).

One important type of independence graph is directed and without cycles. It has at most one directed edge for each node pair, which we call an arrow, and it is acyclic because we cannot start from any one node, follow arrows pointing in the same direction, and return to the starting node. In our context each arrow points from the node of a directly explanatory variable to the node of a response variable. A variable which is a response to one variable and explanatory to another is called intermediate.

*Address for correspondence*: Nanny Wermuth, Department of Mathematical Statistics, Chalmers, Gothenburg University, Eklandagatan 86, 41296 Göteborg, Sweden.
E-mail: nanny.wermuth@math.chalmers.se


1369–7412/04/66687

Often substantive knowledge is sufficiently strong to relate a response of primary interest via a sequence of intermediate single variables to a purely explanatory variable. Then the corresponding independence graph is directed and acyclic. However, sometimes more complex types of independence graphs are of interest which may have several types of edge and possibly more than one edge for a node pair.

We consider a set $N$ of nodes, completely ordered as $N=(1,\ldots,d_N)$ and having node $i$ correspond to random variable $Y_i$. Their joint distribution is said to be generated over a directed and acyclic graph by starting with the marginal distribution at node $d_N$, continuing with the distribution for the variable at node $d_N-1$ given $d_N$, and with conditional univariate distributions at nodes $i$ given $i+1,\ldots,d_N$ up to the distribution of $Y_1$ depending possibly on all previous variables. A direct proper dependence of $Y_i$ is thereby defined to be on only those $Y_j$ of the potentially explanatory variables $Y_{i+1},\ldots,Y_N$ for which an arrow points from node $j$ to node $i$. Such nodes $j$ are called parents of $i$ and form the set $\mathrm{par}_i$. In a condensed notation the joint density $f_N$ is

$$f_N=\prod_i^{d_N} f_{i|i+1,\ldots,d_N}=\prod_i^{d_N} f_{i|\mathrm{par}_i}, \tag{1}$$

and we concentrate here on non-degenerate families of distributions of this type. For such a factorization each missing $ij$-arrow with $i<j$ means conditional independence of $Y_i$ and $Y_j$ given the variables at the parent nodes of node $i$. This is written compactly as $i \perp\!\!\!\perp j|\mathrm{par}_i$.

The collection of all missing edges is equivalent to a set of independence statements. It defines the independence structure of the graph, i.e. the set of all independences satisfied by all distributions generated over the graph. A directed and acyclic graph with a complete ordering of the nodes is called the parent graph. It prescribes a stepwise process for generating the distribution. We are concerned in this paper with implications of such generating processes when conditioning sets of variable pairs are changed to those specified by different types of chain graph models.

For chain graphs the nodes are arranged in a sequence of $d_{\mathrm{CC}}$ chain components $g$, each containing one or more nodes, so that $N=(1,\ldots,g,\ldots,d_{\mathrm{CC}})$ and the joint density is reconsidered in the form

$$f_N=\prod_{g=1}^{d_{\mathrm{CC}}} f_{g|g+1,\ldots,d_{\mathrm{CC}}}. \tag{2}$$

Within this broad formulation of chain graphs there are two main different possibilities which we call multivariate regression chains and blocked concentration chains.

In graphs of multivariate regression chains all edges are shown in a broken fashion; we call them dashed edges. For a given chain component $g$, variables at nodes $i$ and $j$ are considered conditionally given all variables in chain components $g+1,\ldots,d_{\mathrm{CC}}$. Thus, the univariate and bivariate densities

$$\begin{aligned} &f_{i|g+1,\ldots,d_{\mathrm{CC}}},\\ &f_{ij|g+1,\ldots,d_{\mathrm{CC}}} \end{aligned} \tag{3}$$

determine the presence or absence of a dashed $ij$-arrow, which points to node $i$ in chain component $g$ from a node $j$ in $g+1,\ldots,d_{\mathrm{CC}}$, and of a dashed $ij$-line within $g$ when $j$ itself is in $g$. Accordingly, the meaning of a missing dashed $ij$-arrow and of a missing dashed $ij$-line is

$$\begin{aligned} i \perp\!\!\!\perp j \mid \{g+1,\ldots,d_{CC}\}\setminus j, \\ i \perp\!\!\!\perp j \mid g+1,\ldots,d_{CC}. \end{aligned} \tag{4}$$

In graphs of blocked concentration chains all edges are shown in a solid fashion; we call them full edges. For a given chain component $g$, variables at nodes $i$ and $j$ are considered conditionally given all other variables in $g$ and the variables in $g+1,\ldots,d_{CC}$. Thus, the univariate and bivariate densities

$$\begin{aligned} f_{i\mid g\setminus\{i\},g+1,\ldots,d_{CC}}, \\ f_{ij\mid g\setminus\{i,j\},g+1,\ldots,d_{CC}} \end{aligned} \tag{5}$$

are relevant for a full $ij$-arrow pointing to node $i$ in $g$ and for a full $ij$-line within $g$. Accordingly, the meaning of both a missing full $ij$-arrow and a missing full $ij$-line is

$$i \perp\!\!\!\perp j \mid \{g, g+1,\ldots,d_{CC}\}\setminus\{i,j\}. \tag{6}$$

Chain graphs which we draw with full edges, dashed edges and with mixtures of full lines and dashed arrows have been studied by many researchers; see for example Lauritzen and Wermuth (1989), Frydenberg (1990), Wermuth (1992) and Studený and Bouckaert (1998) for the first type, Cox and Wermuth (1993, 1996) and Richardson (2003) for the second type, Levitz *et al.* (2001) for a mixture and Wermuth and Cox (2001) for these three types.

One central theme of this paper is the examination of the set of independences that results when, starting with a triangular system, the comparison with an analysis is contemplated in which roles of some explanatory and some response variables may be interchanged and joint responses are permitted. Put differently, we derive chain graphs which are induced, i.e. implied, by the parent graph of a given triangular system, whenever some subsets of variables are reconsidered, possibly jointly and with different, fixed conditioning sets. In the induced chain graph an edge is missing if and only if the corresponding independence statement is implied by the generating process (1). Such reconsiderations arise typically when there is some uncertainty about the actual ordering of the variables or when there are alternative substantive hypotheses (Wermuth and Lauritzen, 1990) for the variables under study.

This is illustrated here with the following small example from interdisciplinary research on diabetes. Physicians and psychologists searched for important determinants of whether patients succeed well in controlling their chronic disease. The data are reproduced in the appendix of Cox and Wermuth (1996). For each of two patient groups with different levels of formal schooling the parent graph that is shown in Fig. 1 fits the data well. It postulates a stepwise process by which the data could have been generated, corresponding to the ordering $N=(Y,X,Z,W)$ and to the factorization of the joint density (1) read directly off the graph as

$$f_N = f_{Y\mid XW}\, f_{X\mid Z}\, f_{Z\mid W}\, f_W.$$

It might now be argued that acquiring knowledge about diabetes and controlling the blood sugar are both consequences of the illness or that having high blood sugar levels leads a patient to learning more about the illness. This questions the above ordering of the variables and we might ask: supposing that the generating process of Fig. 1 is correct, what are the consequences

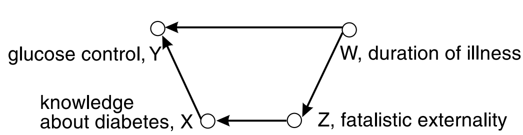


**Fig. 1.** Postulated generating graph for data on glucose control

for analysing the variables $Y$ and $X$ as responses on an equal footing or for regarding $X$ as a response to $Y$?

Such questions concerning relationships that are induced by a given parent graph for a prespecified chain graph may be answered in several ways. One is to use the factorization of the joint density as implied by the generating graph and to find directly whether it induces a factorization in the relevant densities (3) or (5). Another is by tracing special paths, i.e. by finding sequences of adjacent edges in the generating graph identified by a separation criterion. Such criteria specify conditions under which a directed and acyclic graph induces any given conditional independence statement. Different but equivalent criteria have been formulated so far exclusively in terms of paths in graphs (Pearl, 1988; Lauritzen *et al.*, 1990; Wermuth and Cox, 1998a). A further possibility is pursued in the present paper. It is to obtain the information on the presence and absence of the edges induced in a prespecified chain graph by transformations of a matrix form of the generating graph. The approach and formulation that are adopted here are different from other generalizations of directed acyclic graphs analysed by Koster (2002), by Richardson and Spirtes (2002) or by Cox and Wermuth (1996), chapter 8.

For a chain order specified by $N = (\{Y, X\}, \{Z\}, \{W\})$ two chain graphs that are induced by the parent graph of Fig. 1 are shown in Fig. 2. For the multivariate regression chain that is induced by Fig. 1 an additional dashed $YZ$-edge results from marginalizing implicitly over $X$, an intermediate variable between $Y$ and $Z$ in Fig. 1. For the blocked concentration chain an additional full $XW$-edge results from conditioning implicitly on the common response $Y$ of both $X$ and $W$ in Fig. 1.

For the remainder of this paper we use for illustration the parent graph of moderate size that is shown in Fig. 3. All arrows in it point from nodes with larger numbers to nodes with smaller numbers. The matrix form of such a parent graph is upper triangular. Node $i$ corresponds to row $i$ in the matrix that is obtained from an identity matrix of dimension $d_N$ by inserting an $ij$-1, i.e. a 1 in position $(i, j)$ for $i < j$, if and only if there is an $ij$-arrow in the parent graph. We call this matrix the edge matrix of the parent graph or the generating edge matrix.

As one application we show with Fig. 8 in Section 9 the graph of a selected blocked concentration chain as it is induced by the generating graph of Fig. 3. In this case four chain components have been chosen: $N = (a, b, c, C)$ with $a = \{7, 12, 14\}$, $b = \{1, 4, 11, 13\}$, $c = \{2, 8, 9\}$, $C = \{3, 5, 6, 10\}$ and $C$ playing the role of a general conditioning variable. This choice of chain

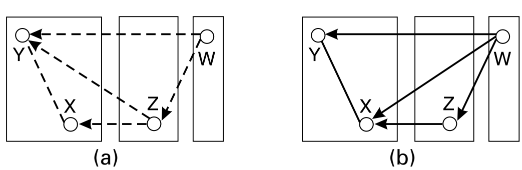


**Fig. 2.** Chain graphs induced by the graph of Fig. 1 for $N = (\{Y, X\}, \{Z\}, \{W\})$ in (a) a multivariate regression chain and in (b) a blocked concentration chain

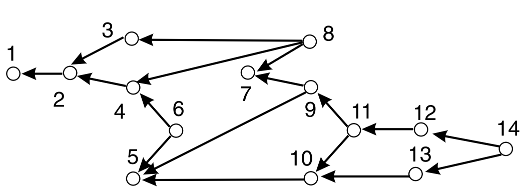


**Fig. 3.** Generating graph in 14 nodes

components requires from expression (2) the joint density to be reconsidered in the form

$$f_N = f_{a|b,c,C}\, f_{b|c,C}\, f_{b|C}\, f_C.$$

Then the choice of a family of chain graphs fixes the conditioning set for each variable pair.

The new factorization shows that in induced chain graphs it is simple to condition on all variables at nodes in the last chain component $C$, or to marginalize over all variables in the first component. This is especially important whenever two investigations and analyses are to be compared which agree on a core set of variables but differ somewhat, for example because in one study there are no observations on a subset of variables or only a subpopulation corresponding to fixed level combination of another subset of variables is investigated.

### *1.2. Outline of the paper*

The paper starts in Section 2 by distinguishing a number of types of graph that are central to our discussion of graphs that are induced by triangular systems (1). An edge in such a graph has for Gaussian random variables a direct interpretation as a particular linear association.

Linear triangular systems, for which a full distributional specification is not needed, are introduced in Section 3 and results are developed in Sections 4–6 which lead to induced linear associations in covariance, concentration and regression coefficient matrices. One key idea is that of orthogonalizing relationships between vector variables so that, when one component is held fixed, the relationships of the other component remain unaffected.

To link the results for linear triangular systems to those for arbitrary distributions generated over the same parent graph, we introduce edge matrix forms of graphs, i.e. matrices of 0s and 1s, with 0s specifying missing edges in induced graphs and, at the same time, so-called structural 0s in induced linear associations.

In Section 7 the connection to arguments based directly on the factorization of densities is shown. The interplay between these and edge inducing paths in graphs derived from edge matrices is the basis for extending the results for linear systems to arbitrary distributions. In Section 8 the general results for induced chain graph models are given. Some ways of using the results are illustrated in Section 9.

## 2. Graph terminology and special induced chain graphs

### *2.1. Types of nodes and paths in parent graphs*

A subgraph that is induced by a subset of nodes in a given graph consists of the nodes in the subset and of the edges among them in the graph. In a directed and acyclic graph there are three different types of subgraphs in three nodes having two edges and called V-configurations:

$$i \leftarrow j \leftarrow k, \qquad j \leftarrow l \rightarrow k, \qquad i \rightarrow h \leftarrow j.$$

We use throughout the convention that $h < i < j < k < l$. A node that is connected to another by an edge is called a neighbour. In the above V-configurations a common neighbour of two nodes is either a transition node, having an incoming and an outgoing arrow (left), a common source node, having two outgoing arrows (middle) or a common sink or collision node, having two incoming arrows (right). Accordingly, the three types of V-configuration are said to be transition oriented, source oriented and sink oriented. It has also become a convention to say, interchangeably, for the left-hand case that node $i$ is a descendant of $j$ and $k$, but a child only of $j$ and that $k$ and $j$ are ancestors of $i$, but only $j$ is a parent of $i$. A path is a sequence of adjacent edges, irrespective of their orientation. A path of arrows leaving from $l$ and leading to $i$, with only transition nodes along it, is a direction preserving path with $l$ being an ancestor of $i$. Node

$l$ is a common ancestor of nodes $i$ and $j$ if a direction preserving path leads from $l$ to both $i$ and $j$. Any given node may play different roles with respect to different neighbours.

The statistical meaning of variables corresponding to the different types of node is as follows. The variable at a parent node is directly explanatory for the variable at a child node. The variable at an ancestor but not parent node is indirectly explanatory for the variable at the descendant node. The variable at a common ancestor node is a common explanatory variable. Further the variable at a transition node is an intermediate variable, at a common sink node it is a common response and at a common source node it is a common directly explanatory variable.

Given this interpretation it is almost immediate that for V-configurations marginalizing over a transition node or over a common source node and conditioning on a common sink node are all edge inducing. An induced edge in an independence graph means in general that a specific independence statement in the parent graph no longer holds in all distributions that are generated over the new graph. In particular, an association may become induced for some such distributions after ignoring an important intermediate variable or a common directly explanatory variable and after selecting a subpopulation defined by the levels of a common response. Induced edges may correspond to strong induced associations. For examples see Wermuth and Cox (1998b) and Wermuth (2003).

We now summarize the types of induced chain graph, i.e. graphs implied by the generating graph. From expression (2), in particular, a chain graph with a single chain component is an undirected graph and an ordering in which each chain component contains a single node gives again a graph which is directed and acyclic.

To describe different types of induced systems we proceed in a few steps. First we define undirected graphs for the node set $N$. We then suppose that the nodes of $N$ are divided into two sets $a$ and $b$ such that the variables in set $a$ are regarded as being on an equal footing and as responses to the variables in $b$. This leads to considering conditional densities $f_{a|b}$ as joint response regressions and to calling their independence graphs joint response regression graphs. Finally we view chain graphs as an ordered sequence of joint response regression graphs.

### 2.2. Overall concentration and covariance graphs

We consider two types of undirected graph. Both have at most one edge for each node pair but differ in the meaning of their edges. If defined for the overall node set $N$ the undirected graph with full, i.e. solid, lines is called the overall concentration graph. It concerns the conditional relations of each pair given all remaining variables of $N$. We use throughout relations as the generic term for both independences and associations.

In contrast, an undirected graph that is defined for node set $N$ with dashed, i.e. broken, lines is called a covariance graph. It concerns the marginal pairwise relations. The dashed lines remind us that the conditioning set for each edge, be it present or missing, is smaller in a covariance graph than in the full line graph. The conditional independence statements attached to a missing $ij$-edge differ accordingly. We write them again in terms of nodes, as respectively

$$i \perp\!\!\!\perp j | N \backslash \{i, j\}, \qquad i \perp\!\!\!\perp j.$$

### 2.3. Joint response regression graphs

Suppose next that the nodes of $N$ are divided into just two sets, $N = (a, b)$, associated with vector random variables $Y_a$ and $Y_b$. Then the joint response regression graph is an independence graph for the conditional distribution of $Y_a$ given $Y_b$. We represent the conditioning on $Y_b$ in a diagram by enclosing $b$ with a doubly lined box and by not showing any edges within $b$. Any remaining

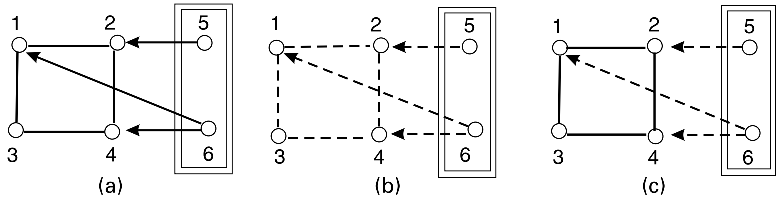


**Fig. 4.** Simple examples of three types of joint regression graphs for $Y_a$ with $a=\{1,2,3,4\}$ given $Y_b$ with $b=\{5,6\}$: graphs of (a) blocked concentrations (block regression), (b) multivariate regressions and (c) concentration regressions, e.g. $2 \perp\!\!\!\perp 6|N\backslash\{2,6\}$ in (a) but $2 \perp\!\!\!\perp 6|5$ in (b)

pair of nodes in $N$ is connected by at most one edge. An edge is undirected if both nodes are in $a$ and directed from $b$ to $a$ if one node is in $b$, and the other in $a$. Three types of such joint response graphs are illustrated in simple form in Fig. 4. A blocked concentration graph, which in the literature has also been named a block regression graph (Fig. 1(a)), has full lines within $a$ and full arrows pointing from $b$ to $a$. Each $ij$-edge concerns a conditional relation of $Y_i$ and $Y_j$ given all remaining variables in $N$. Thus, the independence statement that is attached to each missing $ij$-edge is

$$i \perp\!\!\!\perp j|N\backslash\{i,j\}.$$

A multivariate regression graph (Fig. 4(b)) has dashed lines within $a$ and dashed arrows pointing from nodes in $b$ to nodes in $a$. An $ij$-edge concerns the conditional relation between $Y_i$ and $Y_j$ given all remaining variables in $b$. Thus, the independence statement that is attached to a missing $ij$-edge is

$$i \perp\!\!\!\perp j|b \quad \text{or} \quad i \perp\!\!\!\perp j|b\backslash j,$$

depending on whether $i$ and $j$ are both in $a$, or $i$ is in $a$ and $j$ is in $b$.

A concentration regression graph (Fig. 4(c)) is a mixture of the other two types of regression graph, having full lines within $a$ and dashed arrows pointing from nodes in $b$ to nodes in $a$. Thus, in particular, a missing $ij$-edge means

$$i \perp\!\!\!\perp j|N\backslash\{i,j\} \quad \text{or} \quad i \perp\!\!\!\perp j|b\backslash j,$$

depending on whether $i$ and $j$ are both in $a$, or $i$ is in $a$ and $j$ is in $b$.

The graphs that are formed from full arrows and dashed lines, which are not considered here, may correspond to models with variation-dependent parameter sets even in the linear case.

### *2.4. Chain graphs as sequences of regression graphs*

Chain graphs can now be viewed as recursive sequences of joint response regression graphs to which a marginal covariance or concentration graph has been added for component $d_{\mathrm{CC}}$. For just two chain components the meaning of missing lines within $a$ and missing arrows between $a$ and $b$ is as described above in Section 2.3. A missing edge in the marginal concentration and covariance graph of $b \subset N$ means respectively

$$i \perp\!\!\!\perp j|b\backslash\{i,j\}, \qquad i \perp\!\!\!\perp j.$$

More generally, for a chain that is defined by $N=(1,\ldots,d_{\mathrm{CC}})$, the variables in each chain component $g$ are considered on an equal footing and as potential responses to the explanatory variables in $r=\{g+1,\ldots,d_{\mathrm{CC}}\}$, the relevant joint conditional density being $f_{g|g+1,\ldots,d_{\mathrm{CC}}}=f_{g|r}$.

## 3. Matrix terminology and linear triangular systems

### 3.1. Partitioning and transforming matrices

We write an invertible square matrix $M$ and its inverse $M^{-1}$ partitioned into two components $a$ and $b$ as

$$M = \begin{pmatrix} M_{aa} & M_{ab} \\ M_{ba} & M_{bb} \end{pmatrix},$$

$$M^{-1} = \begin{pmatrix} M^{aa} & M^{ab} \\ M^{ba} & M^{bb} \end{pmatrix}$$

and summarize three standard matrix results as follows.

*Lemma 1* (block triangularization of a square matrix). A square matrix $M$ is block triangularized by premultiplying it by a matrix $T$, where

$$T = \begin{pmatrix} I_{aa} & 0 \\ -M_{ba}M_{aa}^{-1} & I_{bb} \end{pmatrix},$$

$$TM = \begin{pmatrix} M_{aa} & M_{ab} \\ 0 & M_{bb.a} \end{pmatrix},$$

$I$ denotes an identity matrix, $M_{aa}^{-1}$ means throughout $(M_{aa})^{-1}$ and $M_{bb.a} = M_{bb} - M_{ba}M_{aa}^{-1}M_{ab}$.

*Lemma 2* (block diagonalization of a triangular matrix). An upper triangular matrix $N$ is block diagonalized by premultiplying it by a matrix $D$, where

$$D = \begin{pmatrix} I_{aa} & -N_{ab}N_{bb}^{-1} \\ 0 & I_{bb} \end{pmatrix},$$

$$DN = \begin{pmatrix} N_{aa} & 0 \\ 0 & N_{bb} \end{pmatrix}.$$

*Lemma 3* (block diagonalization of a symmetric matrix). A symmetric matrix $\Sigma$ is block diagonalized by premultiplying it by a matrix $P$ and post-multiplying it by the transpose $P^{\mathrm{T}}$ of $P$, where

$$P = \begin{pmatrix} I_{aa} & -\Sigma_{ab}\Sigma_{bb}^{-1} \\ 0 & I_{bb} \end{pmatrix},$$

$$P\Sigma P^{\mathrm{T}} = \begin{pmatrix} \Sigma_{aa.b} & 0 \\ 0 & \Sigma_{bb} \end{pmatrix}.$$

There are two rather different interpretations of the last result. Let $Y$ be a mean-centred random vector variable having covariance matrix $\Sigma$ and concentration matrix $\Sigma^{-1}$ and let $\Pi_{a|b}$ be the matrix of regression coefficients of $Y_b$ in linear least squares regression of $Y_a$ on $Y_b$, having $\beta_{ij.b\setminus j}$ as elements. For an account of linear least squares regression for random variables emphasizing the absence of detailed distributional assumptions, see Cramér (1946), page 302. Then the following expressions of regression coefficients in terms of covariances and in terms of concentrations

$$\Pi_{a|b} = \Sigma_{ab}\Sigma_{bb}^{-1} = -(\Sigma^{aa})^{-1}\Sigma^{ab}$$

result from an equality that is derived for the symmetric partitioned matrices $\Sigma$ and $\Sigma^{-1}$ for which a statistical interpretation has been given by Dempster (1969), pages 113 and 177, namely

$$\begin{pmatrix} \Sigma_{aa.b} & \Sigma_{ab}\Sigma_{bb}^{-1} \\ . & \Sigma_{bb}^{-1} \end{pmatrix} = \begin{pmatrix} (\Sigma^{aa})^{-1} & -(\Sigma^{aa})^{-1}\Sigma^{ab} \\ . & \Sigma^{bb.a} \end{pmatrix}, \tag{7}$$

where $\Sigma^{bb.a} = \Sigma^{bb} - (\Sigma^{ab})^{\mathrm{T}}(\Sigma^{aa})^{-1}\Sigma^{ab}$. Here and throughout we use a dot to denote the lower off-diagonal section of a symmetric matrix. Therefore, the matrix product $PY$ transforms $Y_a$ into $Y_{a|b} = Y_a - \Pi_{a|b}Y_b$ which corresponds for mean-centred joint Gaussian distributions to subtracting from $Y_a$ its conditional mean

$$E(Y_a|Y_b = y_b) = \Pi_{a|b}y_b.$$

Alternatively, we can say that the matrix $P$ orthogonalizes the random vector $Y$ into uncorrelated components $Y_{a|b}$ and $Y_b$ with

$$PY = \begin{pmatrix} I_{aa} & -\Pi_{a|b} \\ 0 & I_{bb} \end{pmatrix} \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} = \begin{pmatrix} Y_{a|b} \\ Y_b \end{pmatrix}.$$

For an ordered partitioning into three components $N = (a, b, c)$ the partitioned forms may be written with $H = (a, b)$ and $K = (b, c)$ as

$$\Sigma_{HH.c} = \begin{pmatrix} \Sigma_{aa.c} & \Sigma_{ab.c} \\ . & \Sigma_{bb.c} \end{pmatrix},$$

$$\Sigma_{KK}^{-1} = \begin{pmatrix} \Sigma^{bb.a} & \Sigma^{bc.a} \\ . & \Sigma^{cc.a} \end{pmatrix}.$$

By using $\Sigma_{bb.c}^{-1} = \Sigma^{bb.a}$ (see for example Dempster (1969)), it follows that

(a) the covariance matrix of $Y_{b|c}$ is the $(b, b)$ submatrix of $\Sigma_{HH.c} = (\Sigma^{HH})^{-1}$, i.e. corresponds to components within $b$,
(b) the concentration matrix of $Y_{b|c}$ is the $(b, b)$ submatrix of $\Sigma_{KK}^{-1} = \Sigma^{KK.a}$ and
(c) the regression coefficient matrix for linear regression of $Y_b$ on $Y_c$ is the $(b, c)$ submatrix of $\Pi_{H|c}$, i.e. rows correspond to components $b$ and columns to components $c$,

$$\begin{aligned} \Sigma_{bb.c} &= [\Sigma_{HH.c}]_{b,\,b}, \\ \Sigma_{bb.c}^{-1} &= [\Sigma^{KK.a}]_{b,\,b}, \\ \Pi_{b|c} &= [\Pi_{H|c}]_{b,\,c}. \end{aligned} \tag{8}$$

Matrix versions of recursion relations for covariances and concentrations are

$$\Sigma_{ab.c} = \Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}, \tag{9}$$

$$\Sigma^{bc.a} = \Sigma^{bc} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ac}. \tag{10}$$

The matrix version of Cochran's (1938) recursion relation among regression coefficients gives the matrix of least squares partial regression coefficients, $\Pi_{a|c.b}$, in the form

$$\Pi_{a|c.b} = \Pi_{a|c} - \Pi_{a|b.c}\Pi_{b|c}. \tag{11}$$

We obtain this from

$$\begin{pmatrix} I_{aa} & -\Pi_{a|b.c} & -\Pi_{a|c.b} \\ 0 & I_{bb} & -\Pi_{b|c} \\ 0 & 0 & I_{cc} \end{pmatrix}^{-1} = \begin{pmatrix} I_{aa} & \Pi_{a|b.c} & \Pi_{a|c} \\ 0 & I_{bb} & \Pi_{b|c} \\ 0 & 0 & I_{cc} \end{pmatrix}.$$

The strength of the linear relation between random variables $Y_i$ and $Y_j$ given $Y_C$ is measured by the partial correlation coefficient $\rho_{ij.C}$, which may be found from any one of $\Sigma_{SS.C}$, $\Sigma^{NN}$ and

$\beta_{ij.C}$, where $N=(S,C)$ and $S=\{(i,j)\}$, as

$$\rho_{ij.C}=\sigma_{ij.C}/(\sigma_{ii.C}\sigma_{jj.C})^{1/2}=-\sigma^{ij}/(\sigma^{ii}\sigma^{jj})^{1/2}=\beta_{ij.C}(\sigma_{jj.C}/\sigma_{ii.C})^{1/2}. \tag{12}$$

For a joint Gaussian distribution the statements $\rho_{ij.C}=0$ and $i \perp\!\!\!\perp j|C$ are equivalent. Therefore, a missing edge in an induced chain graph for Gaussian triangular systems indicates both an independence statement and a structural 0 correlation in a concentration, covariance or regression coefficient matrix. Structural 0s in this context are defined to be zero entries which are determined by a given parent graph, i.e. which hold for all linear equations generated over the same graph, and which do not occur only because of special combinations of parameters and their particular values.

### 3.2. Linear triangular systems

For a linear triangular system which has uncorrelated residuals $\varepsilon_i$ we start with a mean-centred column vector variable $Y$ of dimension $d_N$. A family of models is defined by the equations

$$AY=\varepsilon, \tag{13}$$

where $A$ is a family of upper triangular matrices with 1s along the diagonal, with certain off-diagonal elements specified to be 0, and the remainder to be non-zero. The non-singular covariance matrix of the residuals, $\Delta=\text{diag}(\delta_{ii})$, is a diagonal matrix. In effect, therefore the $i$th row specifies $Y_i$ via a linear least squares regression on $Y_{i+1},\ldots,Y_{d_N}$ with a residual uncorrelated with these latter variables. No special form of distribution is assumed for the residuals. In econometrics equation (13) is known as a system of linear recursive regression equations with uncorrelated errors. The covariance matrix of $Y$ and its inverse, the concentration matrix, are respectively

$$\text{cov}(Y)=\Sigma=A^{-1}\Delta A^{-T},$$
$$\text{con}(Y)=\Sigma^{-1}=A^{\mathrm{T}}\Delta^{-1}A.$$

An element in position $(i,j)$ of $A$ is $-\beta_{ij.i+1,\ldots,j-1,j+1,\ldots,d_N}$, i.e. is minus the regression coefficient of $Y_j$ in linear regression of $Y_i$ on $Y_{i+1},\ldots,Y_{d_N}$. The diagonal elements $\delta_{ii}$ of $\Delta$ are the residual variances, $\delta_{ii}=\sigma_{ii.i+1,\ldots,d_N}$. An element in position $(i,j)$ of $B=A^{-1}$ is $\beta_{ij.j+1,\ldots,d_N}$, the regression coefficient of $Y_j$ in linear regression of $Y_i$ on $Y_j,\ldots,Y_{d_N}$. This follows by inverting the triangular matrix $A$ defining the family (Wermuth and Cox (1998a), appendix 1), and simplifying the results by the recursion relations for regression coefficients.

For an arrangement of $Y$ into an arbitrarily chosen component $a$ and the remaining part $b=N\backslash a$, the corresponding two sets of equations are written in matrix form as

$$\tilde{A}\begin{pmatrix}Y_a\\Y_b\end{pmatrix}=\begin{pmatrix}A_{aa}&\tilde{A}_{ab}\\\tilde{A}_{ba}&A_{bb}\end{pmatrix}\begin{pmatrix}Y_a\\Y_b\end{pmatrix}=\begin{pmatrix}\varepsilon_a\\\varepsilon_b\end{pmatrix}. \tag{14}$$

The matrix $\tilde{A}$ can be expressed as $A$ premultiplied and post-multiplied by a permutation matrix and it is in general not of upper triangular form. The two matrices $\tilde{A}_{ab}$ and $\tilde{A}_{ba}$ have jointly at most $d_a d_b$ non-zero elements, since they arise from the upper triangular matrix $A$ after having changed the ordering of the variables. The identities $\tilde{A}_{aa}\equiv A_{aa}$ and $\tilde{A}_{bb}\equiv A_{bb}$ indicate that the original ordering of indices in $A$ is preserved within subsets $a$ and $b$. For $N=(a,b)$ the conditional concentration matrix of $Y_{a|b}$, $\Sigma_{aa.b}^{-1}$, and the marginal covariance matrix of $Y_b$, $\Sigma_{bb}$, which are submatrices of the overall concentration and covariance matrix, can now be written in terms of parameters of the linear triangular system as

$$\begin{aligned}\Sigma_{aa.b}^{-1}&=\Sigma^{aa}=[A^{\mathrm{T}}\Delta^{-1}A]_{a,a},\\ \Sigma_{bb}&=(\Sigma^{bb.a})^{-1}=[A^{-1}\Delta A^{-T}]_{b,b}.\end{aligned} \tag{15}$$

### 3.3. *Edge matrices of overall graphs induced by a parent graph*

The parent graph of a linear triangular system is defined for node set $N$ by attaching no arrow pointing from node $j$ to node $i$ if and only if $a_{ij}=0$ in the family of matrices $A$. Repeated application of the recursion relation (11) gives the non-zero parameters as

$$\begin{aligned} a_{ij} &= -\beta_{ij.\mathrm{par}_i\setminus j}, \\ \delta_{ii} &= \sigma_{ii.\mathrm{par}_i}. \end{aligned} \tag{16}$$

We now consider three types of graph that are induced by a given parent graph for all nodes: the overall concentration graph, the overall ancestor graph and the overall covariance graph. For a linear triangular system a missing $ij$-edge in each of these induced graphs indicates a structural 0 in the corresponding induced parameter matrix, i.e. a 0 in position $(i, j)$ of $\Sigma^{-1}$, of $B=A^{-1}$ and of $\Sigma$ for every member in the family generated over the given parent graph.

To illustrate the following ideas in simplest form, consider $A_r$ for $r=1,2,3$:

$$\begin{array}{ccc} 2 \rightarrow 1 \leftarrow 3, & 1 \leftarrow 2 \leftarrow 3, & 1 \leftarrow 3 \rightarrow 2, \\ A_1=\begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & A_2=\begin{pmatrix} 1 & a_{12} & 0 \\ 0 & 1 & a_{23} \\ 0 & 0 & 1 \end{pmatrix}, & A_3=\begin{pmatrix} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \\ 0 & 0 & 1 \end{pmatrix}. \end{array}$$

Then a multiple of $a_{12}a_{13}$ is introduced in position (2,3) of $\Sigma^{-1}=A_1^{\mathrm{T}}\Delta^{-1}A_1$, a multiple of $a_{12}a_{23}$ in position (1,3) of $B=A_2^{-1}$ and a multiple of $a_{13}a_{23}$ in position (1,2) of $\Sigma=A_3^{-1}\Delta A_3^{-\mathrm{T}}$. In all three cases the conditioning set of the non-adjacent variable pair in the parent graph is modified by a common neighbour node. For instance, the zero element $a_{23}$ in $A_1$ corresponds to a zero marginal correlation and it leads to a non-zero partial correlation corresponding to position (2,3) of $\Sigma^{-1}$. For instance, in the first example, node 1 is a common sink node and the sink-oriented V-configuration induces an edge for the non-adjacent pair (2,3); it is also association inducing.

We derive induced graphs exclusively via transformations of a matrix form of the generating graph. The definition of a corresponding parameter matrix of a linear system in terms of the parameters $A$ and $\Delta$ provides guidance for the type of transformations that are needed and for the type of paths that induce additional edges.

Let $M$ be any matrix. Then the indicator matrix of $M$, denoted by $\mathrm{In}[M]$, is a matrix of 0s and 1s which has a zero element if and only if the corresponding element of $M$ is 0. The generating edge matrix $\mathcal{A}$ is defined to be the indicator matrix that is associated with the family of matrices $A$,

$$\mathcal{A}=\mathrm{In}[A],$$

i.e. $\mathcal{A}$ has a 0 if and only if every member of the family has a 0, the corresponding edge in the parent graph being missing. Then edge matrices of induced graphs arise, as shown below, as matrices of 0s and 1s by appropriately transforming $\mathcal{A}$. A family of densities may be generated over a parent graph with a given edge matrix $\mathcal{A}$; then this graph defines a corresponding family of matrices $A$ and we use also the notation $\mathcal{A}=\mathrm{Ed}[A]$.

Let $L$ be a parameter matrix, such as a covariance matrix, defined in terms of components of the matrices $A$ and $\Delta$ of a linear triangular system, and let $\mathcal{L}=\mathrm{Ed}[L]$ denote the edge matrix of the corresponding induced graph. Then the following statements are equivalent by definition:

(a) there is a missing $ij$-edge in the induced graph,
(b) there is an $ij$-0 in the edge matrix $\mathcal{L}$ and
(c) there is a structural $ij$-0 in the parameter matrix $L$.

For a generating graph in node set $N$ with edge matrix $\mathcal{A}$, the edge matrix of the induced overall concentration graph is denoted by $\mathcal{S}^{NN}$ and is defined to be $\mathcal{S}^{NN}=\mathrm{Ed}[A^{\mathrm{T}}\Delta^{-1}A]$, of the induced overall ancestor graph it is $\mathcal{B}=\mathrm{Ed}[A^{-1}]$ and of the overall covariance graph it is $\mathcal{S}_{NN}=\mathrm{Ed}[A^{-1}\Delta A^{-T}]$.

*Lemma 4* (edge matrices of overall graphs induced by parent graph). The edge matrices just defined are indicator matrices of the following non-negative matrices:

(a) $\mathcal{S}^{NN}=\mathrm{In}[\mathcal{A}^{\mathrm{T}}\mathcal{A}]$;
(b) $\mathcal{B}=\mathrm{In}[(2I-\mathcal{A})^{-1}]$;
(c) $\mathcal{S}_{NN}=\mathrm{In}[\mathcal{B}\mathcal{B}^{\mathrm{T}}]$.

*Proof.* The results are a direct consequence of the structural 0s that were obtained in the underlying matrix products defined for families of triangular systems and generated over the parent graph with edge matrix $\mathcal{A}$.

For case (a) in the matrix product $H^{\mathrm{T}}H$ for any upper triangular matrix $H$, an element in position $(j,k)$ is of the form $h_{jk}+\Sigma_{i<j}\, h_{ij}h_{ik}$. Applied to the non-negative matrix product $\mathcal{A}^{\mathrm{T}}\mathcal{A}$ this implies that an entry in position $(j,k)$ of the product is 0 if and only if the $jk$-arrow is missing in the parent graph and the non-adjacent node pair $(j,k)$ has no common sink node in the parent graph. The product $\mathcal{A}^{\mathrm{T}}\mathcal{A}$ is edge inducing since for a $jk$-0 in $\mathcal{A}$ a non-zero entry results for the product in position $(j,k)$ whenever the node pair has a common sink node in the parent graph. In the underlying matrix product $A^{\mathrm{T}}\Delta^{-1}A$ multiplication of $A^{\mathrm{T}}$ by the diagonal matrix $\Delta^{-1}$ amounts to rescaling the elements in $A^{\mathrm{T}}$ and does not affect the position of structural 0s and non-zeros. Accordingly, diagonal matrices can be ignored when deriving induced edge matrices.

For case (b), the definitions that are given in Section 3.1 for an element in position $(i,j)$ of $A$ and of $B=A^{-1}$ imply that there is a structural non-zero $ij$-entry in $B$ if and only if $j$ is connected to $i$ by a direction preserving path. Now the non-negative matrix $(\mathcal{A}-I)^r$ counts the number of distinct direction preserving paths of length $r$ between each node pair in the parent graph. Since the longest direction preserving path among $d_N$ nodes has $d_N-1$ edges nothing is added to the sum for $r>d_N-1$. Thus there is a non-zero entry in position $(i,j)$ of $\Sigma_r\,(\mathcal{A}-I)^r$ if and only if node $j$ is an ancestor of node $i$ in the parent graph. The equality $(2I-\mathcal{A})^{-1}=I+\Sigma_r\,(\mathcal{A}-I)^r$ is based on the matrix analogue of the sum of an infinite geometric series; see for example Searle (1966), page 94. Thus case (b) is a matrix formulation for shortening every ancestor–descendant path in the parent graph by an arrow pointing in the same direction.

For case (c), in the matrix product $HH^{\mathrm{T}}$ for any upper triangular matrix $H$ an element in position $(i,j)$ is $h_{ij}+\Sigma_{k>j}\,h_{ik}h_{jk}$. Applied to the non-negative matrix product $\mathcal{B}\mathcal{B}^{\mathrm{T}}$ this implies that an entry in position $(i,j)$ of the product is 0 if and only if the $ij$-arrow is missing in the overall ancestor graph and the non-adjacent node pair $(i,j)$ has no common source node in the overall ancestor graph. An additional $ij$-1 is introduced whenever non-adjacent node pair $(i,j)$ has a common source node in the overall ancestor graph. □

Fig. 5(a) shows a parent graph in node set $N'=\{1,2,3,4,5,6,10\}$, with the three types of overall induced graphs of lemma 4 in Figs 5(b), 5(c) and 5(d). This parent graph is the subgraph that is induced by nodes $N'$ in Fig. 3.

A path condition on the parent graph for an additional $ij$-edge induced in the overall covariance graph results from the path interpretation of the matrices involved in cases (b) and (c) of lemma 4: node $j$ is either an ancestor of $i$ or nodes $i$ and $j$ are connected by a common source path, i.e. there is a node $k$ which is an ancestor both of $i$ and of $j$.

For the three induced graphs of lemma 4 the sum of products of the edge matrices is edge inducing, i.e. edge inducing paths are specified by the form of the sum of products of the edge matrices. These sums of products are also edge preserving, i.e. an $ij$-1 present in one of its matrix components is also present in the sum of products.

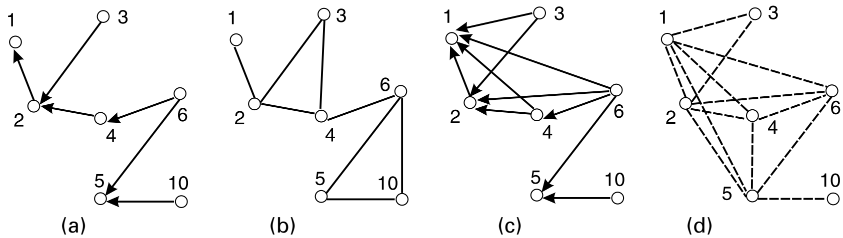


**Fig. 5.** (a) Parent graph for node set $N' = \{1, 2, 3, 4, 5, 6, 10\}$, (b) the overall induced concentration graph, (c) the overall induced ancestor graph and (d) the overall induced covariance graph

For the derivation of edges in an induced graph it is essential that the defining sum of products of edge matrices can be edge inducing and is edge preserving. Thus, the adjacency matrix $(\mathcal{A} - I)$, customarily defined in the graph theory literature as the matrix representation of a directed acyclic graph, is not suitable: it is edge inducing but not edge preserving when premultiplied by its transpose. Similarly, the inverse of a proper edge matrix $(\mathcal{A} \neq I)$ is not used since it contains negative elements and it may also be edge cancelling, such as in

$$\mathcal{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathcal{A}^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix},$$

where the non-zero (1,3) element in $\mathcal{A}$ has become 0 on inversion. Note also that $\mathrm{In}[(2I - \mathcal{B})^{-1}]$ is never edge inducing since every direction preserving path in the parent graph is already closed for $\mathcal{B}$. Thus, matrices of these types are never used when edge matrices for induced graphs are derived; see Sections 3.4 and 6.

### 3.4. Components of induced graphs and stepwise derivations

In this section we derive the matrix formulations of two types of subgraph of the induced overall graphs with edge matrices $\mathcal{S}^{NN}$ and $\mathcal{S}_{NN}$ that were obtained in lemma 4. We also give the matrix formulations of different components of what we call the $a$-line ancestor graph.

We denote by $\mathcal{S}^{aa|b} = \mathrm{Ed}[\Sigma^{-1}_{aa.b}]$ the edge matrix of the subgraph that is induced by nodes $a$ in the overall concentration graph, and by $\mathcal{S}_{bb} = \mathrm{Ed}[\Sigma_{bb}]$ the edge matrix of the subgraph that is induced by nodes $b = N \backslash a$ in the overall covariance graph. We use $\tilde{\mathcal{A}}$ ordered as the underlying matrix $\tilde{A}$ in equation (14) and $\tilde{\mathcal{B}}$ from lemma 4, part (b), being accordingly ordered and partitioned, and we obtain

$$\mathcal{S}^{aa|b} = \mathrm{In}[\mathcal{A}^{\mathrm{T}}_{aa}\mathcal{A}_{aa} + \tilde{\mathcal{A}}^{\mathrm{T}}_{ba}\tilde{\mathcal{A}}_{ba}], \tag{17}$$

$$\mathcal{S}_{bb} = \mathrm{In}[\mathcal{B}_{bb}\mathcal{B}^{\mathrm{T}}_{bb} + \tilde{\mathcal{B}}_{ba}\tilde{\mathcal{B}}^{\mathrm{T}}_{ba}]. \tag{18}$$

These equations are proved by starting from equations (15). It then follows that $\mathcal{S}^{aa|b} = \mathrm{Ed}[A^{\mathrm{T}}\Delta^{-1}A]_{a,a}$. After replacing the parameter matrices by corresponding edge matrices and writing the submatrices explicitly, equation (17) is obtained. The product $\mathcal{A}^{\mathrm{T}}_{aa}\mathcal{A}_{aa}$ is of the same form as the product in lemma 4, part (a), and is thus edge preserving. Also, for every non-adjacent node pair within $a$ in the parent graph an additional edge is induced by $\mathcal{A}^{\mathrm{T}}_{aa}\mathcal{A}_{aa}$ if the pair has a common sink node in $a$ and by $\tilde{\mathcal{A}}^{\mathrm{T}}_{ba}\tilde{\mathcal{A}}_{ba}$ if it has a common sink node in $b$. Hence all relevant edge inducing V-configurations are covered from equation (17). There is a similar type of argument for equation (18) completing the proof.

Next for stepwise evaluation of $\mathcal{B}$ we introduce first the notation $\mathrm{inv}_a(\tilde{A})$ for a generalization of equation (7) to the invertible non-symmetric matrix $\tilde{A}$, which we call its partial inversion on components $a$, and identify components as follows:

$$\mathrm{inv}_a(\tilde{A}) = \begin{pmatrix} A_{aa}^{-1} & \phi_{a|b} \\ \theta_{b|a} & A_{bb.a} \end{pmatrix} = \begin{pmatrix} A_{aa}^{-1} & -A_{aa}^{-1}\tilde{A}_{ab} \\ \tilde{A}_{ba}A_{aa}^{-1} & A_{bb} - \tilde{A}_{ba}A_{aa}^{-1}\tilde{A}_{ab} \end{pmatrix}. \tag{19}$$

Direct matrix computation shows that $\mathrm{inv}_a(\tilde{A})$ turns into $\tilde{B}$, the inverse of $\tilde{A}$, after it has been partially inverted on components $b$. The components of the edge matrix counterpart of equation (19), denoted by $\mathrm{Ed}[\mathrm{inv}_a(\tilde{A})]$, are

$$\mathrm{Ed}[\mathrm{inv}_a(\tilde{A})] = \begin{pmatrix} \mathcal{A}^{aa} & \mathcal{F}_{ab} \\ \mathcal{T}_{ba} & \mathcal{A}_{bb.a} \end{pmatrix} = \begin{pmatrix} \mathrm{In}[(2I_{aa} - \mathcal{A}_{aa})^{-1}] & \mathrm{In}[\mathcal{A}^{aa}\tilde{\mathcal{A}}_{ab}] \\ \mathrm{In}[\tilde{\mathcal{A}}_{ba}\mathcal{A}^{aa}] & \mathrm{In}[\mathcal{A}_{bb} + \mathcal{T}_{ba}\tilde{\mathcal{A}}_{ab}] \end{pmatrix}. \tag{20}$$

To prove these results, we use equation (19) to choose the edge matrix components accordingly for equation (20). Then lemma 4, part (b), is applied to $A_{aa}^{-1}$ and negative signs are ignored as irrelevant for positions of structural 0s. The result then follows from the matrix forms of the sums of products of edge matrices for each component: an edge in $\tilde{A}$ is preserved in $\mathrm{Ed}[\mathrm{inv}_a(\tilde{A})]$ and an additional $ij$-1 arises for a non-adjacent pair $(i, j)$ if and only if $j$ is an $a$-line ancestor of $i$, i.e. an ancestor with all nodes along the ancestor–descendant path in $a \subset N$, thus completing the proof.

With equation (20) every $a$-line ancestor becomes a parent. By turning next every $b$-line ancestor in this graph into a parent, all direction preserving paths are closed, and hence the induced overall ancestor graph results.

To illustrate the construction of the edge matrix of an induced partial ancestor graph we take the subgraph that is induced by $N' = \{1, 2, 3, 4, 5, 6, 8, 10\}$ in Fig. 3 as the parent graph in Fig. 6(a). Fig. 6(b) shows the $a'$-line ancestor graph with $a' = \{2, 6, 8\}$ and Fig. 6(c) shows the overall ancestor graph with $b' = N \backslash a'$ that is obtained as the $b'$-line ancestor graph of Fig. 6(b).

We now use a numerical example to illustrate several issues. Zero off-diagonal elements in $A$ correspond to 0s in the generating family. We take $\tilde{A}_{ba} = 0$, so that

$$\Pi_{a|b} = -A_{aa}^{-1}A_{ab} = -(\Sigma^{aa})^{-1}\Sigma^{ab} = \Sigma_{ab}\Sigma_{bb}^{-1} = B_{ab}B_{bb}^{-1},$$
$$\Sigma_{aa.b} = A_{aa}^{-1}\Delta_{aa}A_{aa}^{-\mathrm{T}},$$
$$\Sigma_{bb}^{-1} = A_{bb}^{\mathrm{T}}\Delta_{bb}^{-1}A_{bb}$$

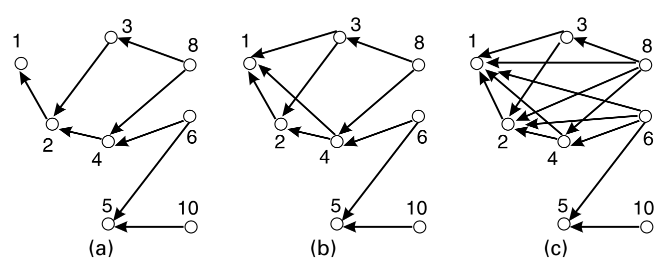


**Fig. 6.** (a) Parent graph with node set $N' = \{1, 2, 3, 4, 5, 6, 8, 10\}$, (b) its $a'$-line ancestor graph, where $a' = \{2, 6, 8\}$, and (c) its overall ancestor graph

holds, $a=(1,2,3)$ and $b=(4,5,6)$, and as numerical values of the parameters we take $\Delta=I$ and

$$A_{aa}=\begin{pmatrix}1&1&-2\\0&1&0\\0&0&1\end{pmatrix},$$

$$A_{ab}=\begin{pmatrix}1&1&0\\1&1&0\\0&0&1\end{pmatrix},$$

$$A_{bb}=\begin{pmatrix}1&1&0\\0&1&1\\0&0&1\end{pmatrix}.$$

Here $\mathcal{A}_{ab}\equiv A_{ab}$ and $\mathcal{A}_{bb}\equiv A_{bb}$. From equations (19) and (20) we obtain $\mathcal{A}^{aa}=\mathrm{Ed}[-A_{aa}^{-1}]\equiv \mathrm{In}[A_{aa}]$ and

$$\mathrm{Ed}[\Pi_{a|b}]=\mathrm{In}[\mathcal{A}^{aa}\mathcal{A}_{ab}]=\begin{pmatrix}1&1&1\\1&1&0\\0&0&1\end{pmatrix},$$

$$\mathrm{In}[\Pi_{a|b}]=\mathrm{In}[-A_{aa}^{-1}A_{ab}]=\begin{pmatrix}0&0&1\\1&1&0\\0&0&1\end{pmatrix}$$

illustrates that with the induced edge matrix it is possible to distinguish between structural 0s and additional 0s which occur only because of special combinations of parameter values.

The example illustrates also that some of the equivalent forms of induced parameter matrices may be unsuitable for deriving structural 0s, since some of their components are never edge inducing. Here, some of the matrix products contain explicitly $B_{bb}^{-1}$ and implicitly $A_{aa}^{-\mathrm{T}}A_{aa}^{\mathrm{T}}$ or $B_{bb}^{-\mathrm{T}}$. There is no transition-oriented V-configuration in the graph with edge matrix $\mathcal{B}_{bb}$ since it is a subgraph of the overall ancestor graph. Hence no additional edge can ever be introduced from it by closing direction preserving paths, i.e. with $\mathrm{In}[(2I_{bb}-\mathcal{B}_{bb})^{-1}]$. Similarly, the product $A_{aa}A_{aa}^{-1}=I_{aa}$, or its transpose, can never lead to an edge inducing path.

In general it may not be possible to deduce from one induced graph which edges are missing in another graph that is induced by the same parent graph. For instance there may be more structural 0s in $\Sigma_{aa.b}=(\Sigma^{aa})^{-1}$ than in $\Sigma^{aa}$. This is the case also in the above numerical example where the conditioning set for pair (2,3) excludes its common sink node 1 in $\Sigma_{aa.b}$ but includes it in $\Sigma^{aa}$. Then the missing edge for pair (2,3) in the induced covariance graph of $(\Sigma^{aa})^{-1}$ cannot be recovered from the induced concentration graph of $\Sigma^{aa}$. There may also be more structural 0s in the induced $\Sigma_{bb}^{-1}$ than in the induced $\Sigma_{bb}$. This is the case in the above numerical example, where the conditioning set for pair (4,6) includes its common transition node 5 in $\Sigma_{bb}^{-1}$ but excludes it in $\Sigma_{bb}$. Then the missing edge in the induced concentration graph of $\Sigma_{bb}^{-1}$ cannot be recovered from the induced covariance graph of $\Sigma_{bb}$. By contrast, as shown in this and the previous section, definitions of the induced parameter matrices directly in terms of components of the generating family of matrices $A$ and $\Delta$ can lead to edge inducing paths and hence to the proper sets of structural 0s and missing edges.

To summarize, let $L$ and $M$ denote two of the induced parameter matrices derived in this paper, such as $-A_{aa}^{-1}$ and $A_{ab}$, for which the edge matrices $\mathcal{L}$ and $\mathcal{M}$ are known, and for which the matrix product $LM$ defines another parameter matrix with a corresponding induced graph of interest. Then the edge matrix of the induced graph is $\mathrm{Ed}[LM]=\mathrm{In}[\mathcal{L}\mathcal{M}]$ provided that this product can be edge inducing and is edge preserving. Similarly, if the sum of two such matrices $L$ and $M$ of the same size defines another such parameter matrix, then $\mathrm{Ed}[L+M]=\mathrm{In}[\mathcal{L}+\mathcal{M}]$ provided that this sum can be edge inducing and is edge preserving, one example being $L=B_{bb}\Delta_{bb}B_{bb}^{\mathrm{T}}$

and $M=\tilde{B}_{ba}\Delta_{aa}\tilde{B}_{ba}^{\mathrm{T}}$ for equation (18). This extends to sums of products containing components in which more than two matrices are relevant for the positions of structural 0s, such as to $A_{bb.a}=A_{bb}-\tilde{A}_{ba}A_{aa}^{-1}\tilde{A}_{ab}$ for equation (20).

Whenever $\tilde{A}_{ba}\neq 0$, the derivation of the induced graph of a regression coefficient matrix, as well as of induced marginal concentration graphs and of induced conditional covariance graphs, becomes more complex. The form of corresponding induced parameter matrices is derived with corollary 1 in equations (23) and (26) in Section 5.1 and extended in applicability with theorem 2 in Section 6, and Section 4 contains the preliminary results that are needed.

## 4. Orthogonalizing general weighted sums of variables

We now consider a vector random variable $Z$ which is specified by the linear system $MZ=\eta$, where $\eta$ is a vector of random terms of zero mean and given covariance matrix; $M$ is a non-singular matrix of constants. Suppose that $Z$ is partitioned into two parts $Z_a$ and $Z_b$ and that it is required to study the marginal distribution of $Z_b$ and of the random variable $Z_{a|b}$ having the conditional distribution of $Z_a$ for a given value of $Z_b$. For this the key idea is the following. We solve in $MZ=\eta$ for $Z_b$ having a residual denoted by $\eta_{b-a}$. We express $Z_a$ in terms of $Z_b$ and a residual denoted by $\eta_{a|b-a}$ and formed to be orthogonal to $\eta_{b-a}$, i.e. to be uncorrelated with it. On conditioning on $Z_b$ this residual is unchanged so we have two separate systems, one for $Z_b$ and one for $Z_a$ given $Z_b$.

Here we wish to express this procedure in a matrix form which

(a) decomposes into three different interpretable steps of transforming $MZ$,
(b) leads to explicit expressions when applied repeatedly to linear triangular systems and
(c) permits us to connect the matrix results for linear triangular systems to transformations of edge matrices of triangular systems of densities.

To achieve this we derive the following key matrix result by using the three basic results in lemmas 1–3.

*Theorem 1* (orthogonalization of weighted sums of variables). Let $Z$ be a column vector of mean-centred random variables having covariance matrix $\omega$ and partitioned into two components $Z_a$ and $Z_b$. Then the weighted sums of $Z_a$ and $Z_b$ specified by $MZ=\eta$, i.e.

$$\begin{pmatrix} M_{aa}Z_a+M_{ab}Z_b \\ M_{ba}Z_a+M_{bb}Z_b \end{pmatrix}=\begin{pmatrix} \eta_a \\ \eta_b \end{pmatrix},$$

are modified into two new uncorrelated systems written as

$$\begin{pmatrix} M_{aa} & 0 \\ 0 & M_{bb.a} \end{pmatrix}\begin{pmatrix} Z_{a|b} \\ Z_b \end{pmatrix}=\begin{pmatrix} \eta_{a|b-a} \\ \eta_{b-a} \end{pmatrix},$$

where $\mathrm{cov}(\eta_{a|b-a},\eta_{b-a})=0$, by taking $(DTMP^{-1})(PZ)=DT\eta$, with

$$DTMP^{-1}=\begin{pmatrix} M_{aa} & 0 \\ 0 & M_{bb.a} \end{pmatrix},$$

$$PZ=\begin{pmatrix} Z_{a|b} \\ Z_b \end{pmatrix},$$

$$DT\eta=\begin{pmatrix} \eta_{a|b-a} \\ \eta_{b-a} \end{pmatrix}.$$

Here $P$ is the matrix which orthogonalizes $Z$ into components $Z_{a|b}$ and $Z_b$, i.e. block diagonalizes their covariance matrix, $T$ is the matrix which block triangularizes $M$ as well as $MP^{-1}$ and $D$ is the matrix which block diagonalizes the matrix $TMP^{-1}$ as well as the covariance matrix $\tau$ of $T\eta$. The matrix $\omega_{ab}\omega_{bb}^{-1}$ of least squares regression coefficients in the linear regression of $Z_a$ on $Z_b$ is related to the weights, i.e. to components of $M$, and to the covariance matrix $\tau$ by

$$\omega_{ab}\omega_{bb}^{-1} = M_{aa}^{-1}(\tau_{ab}\tau_{bb}^{-1})M_{bb.a} - M_{aa}^{-1}M_{ab}. \tag{21}$$

*Proof.* The result follows from lemmas 1–3 with

$$P = \begin{pmatrix} I_{aa} & -\omega_{ab}\omega_{bb}^{-1} \\ 0 & I_{bb} \end{pmatrix},$$

$$T = \begin{pmatrix} I_{aa} & 0 \\ -M_{ba}M_{aa}^{-1} & I_{bb} \end{pmatrix},$$

$$D = \begin{pmatrix} I_{aa} & -\psi_{a|b} \\ 0 & I_{bb} \end{pmatrix},$$

where

$$\psi_{a|b} = (M_{aa}\omega_{ab}\omega_{bb}^{-1} + M_{ab})M_{bb.a}^{-1},$$

and by realizing that the covariance matrix of the residuals $DT\eta$ is block diagonal for non-zero $\tau_{ab}$ if and only if $\psi_{a|b} = \tau_{ab}\tau_{bb}^{-1}$, since

$$D\tau D^{\mathrm{T}} = \begin{pmatrix} \tau_{aa} - \psi_{a|b}\tau_{ba} - \tau_{ab}\psi_{a|b}^{\mathrm{T}} + \psi_{a|b}\tau_{bb}\psi_{a|b}^{\mathrm{T}} & \tau_{ab} - \psi_{a|b}\tau_{bb} \\ . & \tau_{bb} \end{pmatrix}.$$

If $\tau_{ab} = 0$ then

$$0 = \psi_{a|b} = (M_{aa}\omega_{ab}\omega_{bb}^{-1} + M_{ab})M_{bb.a}^{-1}$$

and hence $\omega_{ab}\omega_{bb}^{-1} = -M_{aa}^{-1}M_{ab}$ satisfies equation (21) as required.

## 5. Orthogonalizing two components of a linear triangular system

### 5.1. Uncorrelated linear triangular systems rearranged into two parts

We now apply theorem 1 of Section 4 to two uncorrelated weighted vector variables which arise from a given linear triangular system (13) by reordering the variables as for equation (14).

*Corollary 1* (induced orthogonal linear systems in $Y_{a|b}$ and $Y_b$). Let the uncorrelated weighted sums $A_{aa}Y_a + \tilde{A}_{ab}Y_b$ and $\tilde{A}_{ba}Y_a + A_{bb}Y_b$ be derived from equation (14) by taking the ordering $N = (a, b)$ in the linear triangular system (13) and by using $\tilde{A}Y = \varepsilon$ of equation (14). Let $\tilde{A}$ be block triangularized by $T$. Then the two induced uncorrelated systems in $Y_{a|b}$ and $Y_b$ are

$$\begin{pmatrix} A_{aa} & 0 \\ 0 & A_{bb.a} \end{pmatrix}\begin{pmatrix} Y_{a|b} \\ Y_b \end{pmatrix} = \begin{pmatrix} \varepsilon_{a|b-a} \\ \varepsilon_{b-a} \end{pmatrix} = \begin{pmatrix} \varepsilon_a - \psi_{a|b}\varepsilon_{b-a} \\ \varepsilon_b - \theta_{b|a}\varepsilon_a \end{pmatrix}, \tag{22}$$

where $\mathrm{cov}(\varepsilon_{a|b-a}, \varepsilon_{b-a}) = 0$, but residuals within the two components may be correlated. The new residuals and $\Pi_{a|b}$ in $Y_{a|b} = Y_a - \Pi_{a|b}Y_b$ are given in terms of the appropriate $M$ and $\tau = \mathrm{cov}(T\varepsilon)$, where

$$\tau = \mathrm{cov}(T\varepsilon) = T\tilde{\Delta}T^{\mathrm{T}} = \begin{pmatrix} \Delta_{aa} & -\Delta_{aa}\theta_{b|a}^{\mathrm{T}} \\ . & \Delta_{bb-a} \end{pmatrix},$$

$$
\begin{aligned}
\theta_{b|a} &= \tilde{A}_{ba} A_{aa}^{-1},\\
\phi_{a|b} &= -A_{aa}^{-1} \tilde{A}_{ab},\\
\psi_{a|b} &= -\Delta_{aa} \theta_{b|a}^{\mathrm{T}} \Delta_{bb-a}^{-1},\\
\Pi_{a|b} &= -A_{aa}^{-1} \Delta_{aa} \theta_{b|a}^{\mathrm{T}} \Delta_{bb-a}^{-1} A_{bb.a} + \phi_{a|b}.
\end{aligned} \tag{23}
$$

*Proof.* This results from theorem 1 with the appropriate form of $(DT\tilde{A}P^{-1})(PY) = DT\varepsilon$. □

The associations within the two sets of uncorrelated residuals in $DT\varepsilon$ are most compactly given by the covariance matrix of $\varepsilon_{b-a}$ and by the concentration matrix of $\varepsilon_{a|b-a}$ as

$$\operatorname{cov}(\varepsilon_{b-a}) = \Delta_{bb-a} = \Delta_{bb} + \theta_{b|a} \Delta_{aa} \theta_{b|a}^{\mathrm{T}} = \tau_{bb}, \tag{24}$$

$$\operatorname{con}(\varepsilon_{a|b-a}) = \Delta^{aa+b} = \Delta_{aa}^{-1} + \theta_{b|a}^{\mathrm{T}} \Delta_{bb}^{-1} \theta_{b|a} = \tau_{aa.b}^{-1}, \tag{25}$$

where $\tau_{aa.b} = \tau_{aa} - \tau_{ab} \tau_{bb}^{-1} \tau_{ba}$ denotes the residual covariance of $\varepsilon_a$ after linear least squares regression on $\varepsilon_{b-a}$. The notation $\Delta_{bb-a}$ is chosen to remind us that it is the covariance matrix of residuals for the variable in the margin, $Y_b$, and the notation $\Delta^{aa+b}$ to remind us that it is the concentration matrix of residuals for the variable $Y_a$ considered conditionally given $Y_b = y_b$.

By using corollary 1 and results (24) and (25) the induced parameter matrices $\Sigma^{aa}$ and $\Sigma_{bb}$ in equations (15) can now be expressed as

$$
\begin{aligned}
\Sigma^{aa} &= \Sigma_{aa.b}^{-1} = A_{aa}^{\mathrm{T}} \Delta^{aa+b} A_{aa},\\
\Sigma_{bb} &= (\Sigma^{bb.a})^{-1} = A_{bb.a}^{-1} \Delta_{bb-a} A_{bb.a}^{-\mathrm{T}},
\end{aligned} \tag{26}
$$

so that they and their inverses can be readily interpreted in terms of parameters that are derived from the linear triangular system and can be used to compute the strength of induced linear associations as prescribed in equation (12) at the end of Section 3.1.

### 5.2. *Correlated linear triangular systems rearranged into two parts*

We now apply theorem 1 to two correlated weighted vector variables which arise after having marginalized in the linear triangular system (13) over $a$. This amounts to splitting the node set of a triangular system repeatedly, first as $N = (a, K)$ and then with $K = (b, c)$. We take the linear system that is given by the second of equations (22) in a mean-centred column vector variable $Y_K$ and let the original ordering be preserved within $K$. Then we can write

$$A_{KK.a} Y_K = \varepsilon_{K-a}, \tag{27}$$

where $A_{KK.a}$ is an upper triangular matrix with 1s along the diagonal and the covariance matrix of the residuals $\varepsilon_{K-a}$, denoted by $\Delta_{KK-a}$, is in partitioned form

$$\operatorname{cov}(\varepsilon_{K-a}) = \Delta_{KK-a} = \begin{pmatrix} \Delta_{bb-a} & \theta_{b|a} \Delta_{aa} \theta_{c|a}^{\mathrm{T}} \\ . & \Delta_{cc-a} \end{pmatrix}.$$

For an arrangement of $Y_K$ into component $b$ and the remaining part $c = K \backslash b$, the corresponding two sets of equations are written in matrix form as

$$\tilde{A}_{KK.a} \begin{pmatrix} Y_b \\ Y_c \end{pmatrix} = \begin{pmatrix} A_{bb.a} & \tilde{A}_{bc.a} \\ \tilde{A}_{cb.a} & A_{cc.a} \end{pmatrix} \begin{pmatrix} Y_b \\ Y_c \end{pmatrix} = \begin{pmatrix} \varepsilon_b - \theta_{b|a} \varepsilon_a \\ \varepsilon_c - \theta_{c|a} \varepsilon_a \end{pmatrix} = \begin{pmatrix} \varepsilon_{b-a} \\ \varepsilon_{c-a} \end{pmatrix}. \tag{28}$$

Here the two matrices $\tilde{A}_{bc.a}$ and $\tilde{A}_{cb.a}$ have jointly at most $d_b d_c$ non-zero elements, since they arise from the upper triangular matrix $A_{KK.a}$ after having changed the ordering of the variables. The two weighted sums are orthogonalized as follows.

*Corollary 2* (induced orthogonal linear systems in $Y_{b|c}$ and $Y_c$). Let the correlated weighted sums $A_{bb.a}Y_b + \tilde{A}_{bc.a}Y_c$ and $\tilde{A}_{cb.a}Y_b + A_{cc.a}Y_c$ result from taking $N=(a,b,c)$, $H=(a,b)$ and $K=(b,c)$ in a linear triangular system (13), after having marginalized over $a$. Let $\tilde{A}_{KK.a}$ be block triangularized by $T$. Then the two induced uncorrelated systems in $Y_{b|c}$ and $Y_c$ are

$$\begin{pmatrix} A_{bb.a} & 0 \\ 0 & A_{cc.H} \end{pmatrix} \begin{pmatrix} Y_{b|c} \\ Y_c \end{pmatrix} = \begin{pmatrix} \varepsilon_{b-a|c-H} \\ \varepsilon_{c-H} \end{pmatrix} = \begin{pmatrix} \varepsilon_{b-a} - \psi_{b|c}\varepsilon_{c-H} \\ \varepsilon_c - \theta_{c|H}\varepsilon_H \end{pmatrix}, \tag{29}$$

where $\text{cov}(\varepsilon_{b-a|c-H}, \varepsilon_{c-H}) = 0$, but residuals within the two components may be correlated. The new residuals and $\Pi_{b|c}$ in $Y_{b|c} = Y_b - \Pi_{b|c}Y_c$ are given in terms of the appropriate $M$ and $\tau = \text{cov}(T\varepsilon_{K-a})$, where

$$\tau = T\,\text{cov}(\varepsilon_{K-a})T^{\mathrm{T}} = \begin{pmatrix} \Delta_{bb-a} & \theta_{b|a}\Delta_{aa}\theta_{c|a}^{\mathrm{T}} - \Delta_{bb-a}\theta_{c|b.a}^{\mathrm{T}} \\ \cdot & \Delta_{cc-H} \end{pmatrix},$$

$$A_{cc.H} = A_{cc} - \tilde{A}_{cH}A_{HH}^{-1}\tilde{A}_{Hc} = A_{cc.a} - \tilde{A}_{cb.a}A_{bb.a}^{-1}\tilde{A}_{bc.a},$$

$$\theta_{c|H} = \tilde{A}_{cH}A_{HH}^{-1} = (\theta_{c|a.b}, \theta_{c|b.a}) = (\theta_{c|a} - \theta_{c|b.a}\theta_{b|a}, \tilde{A}_{cb.a}A_{bb.a}^{-1}),$$

$$\psi_{b|c} = (\theta_{b|a}\Delta_{aa}\theta_{c|a}^{\mathrm{T}} - \Delta_{bb-a}\theta_{c|b.a}^{\mathrm{T}})\Delta_{cc-H}^{-1} = \tau_{bc}\tau_{cc}^{-1},$$

$$\Pi_{b|c} = A_{bb.a}^{-1}(\tau_{bc}\tau_{cc}^{-1})A_{cc.H} - A_{bb.a}^{-1}\tilde{A}_{bc.a}.$$

*Proof.* This results from theorem 1 with

$$P = \begin{pmatrix} I_{bb} & -\Pi_{b|c} \\ 0 & I_{cc} \end{pmatrix},$$

$$T = \begin{pmatrix} I_{bb} & 0 \\ -\tilde{A}_{cb.a}A_{bb.a}^{-1} & I_{cc} \end{pmatrix},$$

$$D = \begin{pmatrix} I_{bb} & -(A_{bb.a}\Pi_{b|c} + \tilde{A}_{bc.a})A_{cc.H}^{-1} \\ 0 & I_{cc} \end{pmatrix}$$

and with $(DT\tilde{A}_{KK.a}P^{-1})(PY_K) = DT\varepsilon_{K-a}$. The covariance matrices of the residuals in the two uncorrelated systems of $A_{bb.a}Y_{b|c}$ and $A_{cc.H}Y_c$ are

$$\text{cov}(\varepsilon_{b-a|c-H}) = \tau_{bb.c},$$
$$\text{cov}(\varepsilon_{c-H}) = \tau_{cc},$$

where $\tau_{bb.c}$ is the residual covariance of $\varepsilon_{b-a}$ after linear least squares regression on $\varepsilon_{c-H}$.

To prove that the recursion relation among $\theta$s holds which permits us to write

$$\begin{aligned} \varepsilon_{c-a} - \theta_{c|b.a}\varepsilon_{b-a} &= \varepsilon_c - \theta_{c|b.a}\varepsilon_b - (\theta_{c|a} - \theta_{c|b.a}\theta_{b|a})\varepsilon_a \\ &= \varepsilon_c - \theta_{c|b.a}\varepsilon_b - \theta_{c|a.b}\varepsilon_a \\ &= \varepsilon_c - \theta_{c|H}\varepsilon_H, \end{aligned}$$

the special form of the inverse of the following triangular matrix may be used:

$$\begin{pmatrix} I_{aa} & 0 & 0 \\ -\theta_{b|a} & I_{bb} & 0 \\ -\theta_{c|a.b} & -\theta_{c|b.a} & I_{cc} \end{pmatrix}^{-1} = \begin{pmatrix} I_{aa} & 0 & 0 \\ \theta_{b|a} & I_{bb} & 0 \\ \theta_{c|a} & \theta_{c|b.a} & I_{cc} \end{pmatrix},$$

which relates $\theta_{c|a.b}$ to $\theta_{c|b.a}$ as

$$\theta_{c|a.b} = \theta_{c|a} - \theta_{c|b.a}\theta_{b|a},$$

an expression that is analogous to the recursion relation for regression coefficient matrices (11) that was derived in Section 3.1. □

Thus, the parameters in equation (29) have all been expressed in terms of those in the starting system $A_{KK.a}Y_K = \varepsilon_{K-a}$. The induced covariance and concentration matrices of $Y_{b|c}$ are

$$\Sigma_{bb.c} = A_{bb.a}^{-1}\tau_{bb.c}A_{bb.a}^{-\mathrm{T}},$$
$$\Sigma_{bb.c}^{-1} = A_{bb.a}^{\mathrm{T}}\tau_{bb.c}^{-1}A_{bb.a}.$$

## 6. Orthogonalizing three components of a linear triangular system

### 6.1. Parameters in induced orthogonal systems

We now extend the results for induced systems to any three disjoint components $a, b$ and $c$ of $N$ in such a way that all induced parameter matrices are expressed in terms of repeated splits into just two components as given with corollary 1.

*Theorem 2* (induced orthogonal linear systems in three components). For $N = (a, b, c)$ three orthogonal systems are induced by the linear triangular system (13) as

$$\begin{pmatrix} A_{aa} & 0 & 0 \\ 0 & A_{bb.a} & 0 \\ 0 & 0 & A_{cc.ba} \end{pmatrix} \begin{pmatrix} Y_{a|bc} \\ Y_{b|c} \\ Y_c \end{pmatrix} = \begin{pmatrix} \varepsilon_{a|K-a} \\ \varepsilon_{b-a|c-H} \\ \varepsilon_{c-H} \end{pmatrix}. \tag{30}$$

To obtain the induced parameter matrices with $H = (a, b)$ and $K = (b, c)$, we define for example $A_{HH}^{-1}$, $\theta_{c|H}$, $\phi_{H|c}$ and $A_{cc.H}$ by taking $H = a$ and $c = b$, as in the discussion before corollary 1: thus,

(a) for $Y_b|Y_c$,

$$\begin{aligned} \Sigma_{bb.c} &= [A_{HH}^{-1}(\Delta^{HH+c})^{-1}A_{HH}^{-\mathrm{T}}]_{b,b}, \\ \Sigma_{bb.c}^{-1} &= \Sigma^{bb.a} = [A_{KK.a}^{\mathrm{T}}(\Delta_{KK-a})^{-1}A_{KK.a}]_{b,b}, \end{aligned} \tag{31}$$

$$\Pi_{b|c} = [-A_{HH}^{-1}\Delta_{HH}\theta_{c|H}^{\mathrm{T}}(\Delta_{cc-H})^{-1}A_{cc.H} + \phi_{H|c}]_{b,c}; \tag{32}$$

and,

(b) for $Y_a|Y_K$,

$$\Sigma_{aa.K} = A_{aa}^{-1}(\Delta^{aa+K})^{-1}A_{aa}^{-\mathrm{T}},$$
$$\Sigma_{aa.K}^{-1} = \Sigma^{aa} = A_{aa}^{\mathrm{T}}\Delta^{aa+b}A_{aa} = [A^{\mathrm{T}}\Delta^{-1}A]_{a,a},$$
$$\Pi_{a|K} = -A_{aa}^{-1}\Delta_{aa}\theta_{K|a}^{\mathrm{T}}(\Delta_{KK-a})^{-1}A_{KK.a} + \phi_{a|K};$$

and, finally,

(c) for $Y_c$,

$$\Sigma_{cc} = A_{cc.H}^{-1}(\Delta_{cc-H})A_{cc.H}^{-\mathrm{T}} = [B\Delta B^{\mathrm{T}}]_{c,c},$$
$$\Sigma_{cc}^{-1} = \Sigma^{cc.H} = A_{cc.H}^{\mathrm{T}}(\Delta_{cc-H})^{-1}A_{cc.H}.$$

*Proof.* The results follow by combining corollaries 1 and 2 and equations (8). All parameter matrices are either submatrices of the overall covariance or concentration matrix or they are of the form derived from corollary 1 in equations (23) and (26).

### 6.2. Edge matrices for induced orthogonal linear systems

For orthogonal linear systems in two and in three components the edge matrices that are needed additionally to those of lemma 4 are the following three, obtained from equations (23) and (26), and denoted by $\mathcal{S}^{bb} = \mathrm{Ed}[\Sigma_{bb}^{-1}]$, $\mathcal{S}_{aa|b} = \mathrm{Ed}[\Sigma_{aa.b}]$ and $\mathcal{P}_{a|b} = \mathrm{Ed}[\Pi_{a|b}]$:

$$\mathcal{S}^{bb} = \mathrm{Ed}[A_{bb.a}^{\mathrm{T}} \Delta_{bb-a}^{-1} A_{bb.a}],$$
$$\mathcal{S}_{aa|b} = \mathrm{Ed}[A_{aa}^{-1} (\Delta^{aa+b})^{-1} A_{aa}^{-\mathrm{T}}],$$
$$\mathcal{P}_{a|b} = \mathrm{Ed}[-A_{aa}^{-1} \Delta_{aa} \theta_{b|a}^{\mathrm{T}} \Delta_{bb-a}^{-1} A_{bb.a} + \phi_{a|b}].$$

To derive them in terms of indicator matrices we need some preliminary results for $\mathcal{D}_{bb-a} = \mathrm{Ed}[\Delta_{bb-a}]$ and $\mathcal{D}^{aa+b} = \mathrm{Ed}[\Delta^{aa+b}]$. From equations (24) and (25) they are defined and from equation (20) they are obtained in terms of the indicator matrix $\mathcal{T}_{ba}$, which is one of the parts of $\mathrm{Ed}[\mathrm{inv}_a(\tilde{A})]$, as

$$\mathcal{D}_{bb-a} = \mathrm{Ed}[\Delta_{bb} + \theta_{b|a} \Delta_{aa} \theta_{b|a}^{\mathrm{T}}] = \mathrm{In}[I + \mathcal{T}_{ba} \mathcal{T}_{ba}^{\mathrm{T}}], \tag{33}$$

$$\mathcal{D}^{aa+b} = \mathrm{Ed}[\Delta_{aa}^{-1} + \theta_{b|a}^{\mathrm{T}} \Delta_{bb}^{-1} \theta_{b|a}] = \mathrm{In}[I + \mathcal{T}_{ba}^{\mathrm{T}} \mathcal{T}_{ba}]. \tag{34}$$

To find whether there is an $ij$-edge in the graph with edge matrix $\mathcal{D}_{bb-a}$ we interpret again the appropriate matrix products. An edge is required for families of models if and only if nodes $i_b$ and $j_b$, two nodes within $b$, have a common source node in $a$ in the $a$-line ancestor graph. Since $\Delta_{bb-a}$ is the covariance matrix of residuals $\varepsilon_{b-a}$ it is the edge matrix of an induced dashed line graph. Similarly, there is an $ij$-edge in the graph with edge matrix $\mathcal{D}^{aa+b}$ if and only if nodes $i_a$ and $j_a$ have a common sink node in $b$ in the $a$-line ancestor graph. Since $\Delta^{aa+b}$ is the concentration matrix of residuals $\varepsilon_{a|b-a}$ it is the edge matrix of an induced full line graph.

As a further preliminary result we need $\mathrm{Ed}[\Delta_{bb-a}^{-1}]$ and $\mathrm{Ed}[(\Delta^{aa+b})^{-1}]$. On inverting the covariance matrix $\Delta_{bb-a}$ the conditioning set for the covariance of pair $(i, j)$ is increased by all remaining nodes in $b$. On inverting the concentration matrix $\Delta^{aa+b}$ the conditioning set for the concentration of pair $(i, j)$ is decreased by all remaining nodes in $a$. By these changes none of the configurations that generated edges in these graphs is removed; instead every connected subgraph is turned into a complete graph.

Let $\mathcal{S}$ denote the edge matrix of an undirected graph for which every V-configuration becomes edge inducing whenever the underlying symmetric matrix $S$ is inverted. Then the edge matrix of $S^{-1}$, denoted by $\mathrm{clos}(\mathcal{S}) = \mathrm{Ed}[S^{-1}]$, is obtained by closing every V-configuration in $\mathcal{S}$. This leads to $\mathrm{Ed}[(\Delta^{aa+b})^{-1}] = \mathrm{clos}(\mathcal{D}^{aa+b})$, arising from the closing of sink-generated, full line paths, and to $\mathrm{Ed}[\Delta_{bb-a}^{-1}] = \mathrm{clos}(\mathcal{D}_{bb-a})$, arising from the closing of source-generated, dashed line paths.

For a matrix formulation to obtain $\mathrm{clos}(\mathcal{S})$ of an undirected graph we use $\mathcal{S}_1 = \mathrm{In}[\mathrm{triu}(\mathcal{S})]$, where $\mathrm{triu}(\mathcal{S})$ is the upper triangular part of $\mathcal{S}$ including the diagonal, and we interpret $\mathcal{S}_1$ as an edge matrix of a directed graph. Then all three types of V-configurations are closed by computing $\mathcal{S}_2 = \mathrm{In}[\mathcal{S}_1^{\mathrm{T}} \mathcal{S}_1]$, $\mathcal{S}_3 = \mathrm{In}[\{2I - \mathrm{triu}(\mathcal{S}_2)\}^{-1}]$ and $\mathrm{clos}(\mathcal{S}) = \mathrm{In}[\mathcal{S}_3 \mathcal{S}_3^{\mathrm{T}}]$. This follows from lemma 4.

*Theorem 3* (induced edge matrices of orthogonal linear systems). Let $a \subset N$ be an arbitrarily chosen set $a$ and $b = N \backslash a$. Let the parameter matrices of the orthogonalized linear system derived from equation (13) either be split into two components (corollary 1) or be split into three components (theorem 2). Then the following edge matrices specify, together with those of lemma 4, their structural 0s:

(a) $\mathcal{S}^{bb} = \mathrm{In}[\mathcal{A}_{bb.a}^{\mathrm{T}} \operatorname{clos}(\mathcal{D}_{bb-a})\mathcal{A}_{bb.a})]$,
(b) $\mathcal{S}_{aa|b} = \mathrm{In}[\mathcal{A}^{aa} \operatorname{clos}(\mathcal{D}^{aa+b})(\mathcal{A}^{aa})^{\mathrm{T}}]$,
(c) $\mathcal{P}_{a|b} = \mathrm{In}[\mathcal{A}^{aa}\mathcal{T}_{ba}^{\mathrm{T}} \operatorname{clos}(\mathcal{D}_{bb-a})\mathcal{A}_{bb.a} + \mathcal{F}_{ab}]$,

where definitions of the edge matrix given above in equations (20), (33), (34) and of $\operatorname{clos}(\mathcal{S})$ are used.

*Proof.* The result follows from the calculations and properties of induced edge matrices that were given in Sections 3.3 and 3.4, from the additional matrix and edge matrix results that were summarized above in this section and because the sums of products are edge preserving and edge inducing. □

Note that for a matrix product of the type $\Sigma_{i,h}\, a_{ji} b_{ih} a_{hk}$, which defines $\mathcal{S}^{bb}$ in case (a), there is an additional $j_b k_b$-1 if and only if the non-adjacent nodes $j_b$ and $k_b$ are connected by the following edge inducing path or by one of the V-configurations derived from it:

$$j_b \rightarrow i_b — h_b \leftarrow k_b.$$

This path shows up in the $a$-line partial ancestor graph having edge matrix $\mathrm{Ed}[\mathrm{inv}_a(\tilde{A})]$ which is given in equation (20) to which is appended the full line graph with edge matrix $\operatorname{clos}(\mathcal{D}_{bb-a})$.

Similarly it follows for case (b) from the matrix product which defines $S_{aa|b}$ that an additional $i_a h_a$-1 is induced if and only if the non-adjacent nodes $i_a$ and $h_a$ are connected by the following edge inducing path or by one of the V-configurations derived from it:

$$i_a \leftarrow j_a \text{- - -} k_a \rightarrow h_a.$$

This path shows up in the $a$-line partial ancestor graph, with edge matrix $\mathrm{Ed}[\mathrm{inv}_a(\tilde{A})]$ which is given in equation (20) appended by the dashed line graph with edge matrix $\operatorname{clos}(\mathcal{D}^{aa+b})$.

For the induced edge matrix $\mathcal{P}_{a|b}$ in case (c) the graphs having edge matrices $\mathrm{inv}_a(\tilde{\mathcal{A}})$ and $\operatorname{clos}(\mathcal{D}_{bb-a})$ are combined first. It then follows from the form of $\mathcal{P}_{a|b}$ that compared with the edge matrices $\mathcal{F}_{ab}$ and $\mathcal{T}_{ba}$ there is an additional $i_a l_b$-1 in $\mathcal{P}_{a|b}$ if and only if the non-adjacent nodes $i_a$ and $l_b$ are connected by the following edge inducing path or by one of the three- or two-edge paths derived from it that lead from a node in $b$ to a node in $a$:

$$i_a \leftarrow k_a \rightarrow j_b — h_b \leftarrow l_b.$$

The construction of the edge matrix for the conditional covariance graph of $Y_{a|b}$ is illustrated in Fig. 7. With a split of the node set $N' = \{1, 2, 3, 4, 5, 6, 8, 10\}$ of the parent graph of Fig. 6(a) into two components $N' = (a', b')$, with $a' = \{2, 6, 8\}$ and $b' = \{1, 3, 4, 5, 10\}$, the edge matrix $\mathcal{S}^{b'b'}$ of the induced concentration graph of $Y_{b'}$ is obtained with a matrix product as $\mathcal{S}^{b'b'} = \mathrm{In}[\mathcal{A}_{b'b'.a'}^{\mathrm{T}} \operatorname{clos}(\mathcal{D}_{b'b'-a'})\mathcal{A}_{b'b'.a'}]$, where

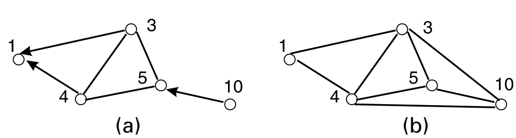


**Fig. 7.** (a) Subgraph induced by $b' = \{1, 3, 4, 5, 10\}$ in the $a'$-line ancestor graph of Fig. 6(b) combined with the source-generated concentration graph with edge matrix $\operatorname{clos}(\mathcal{D}_{b'b'-a'})$ and (b) induced marginal concentration graph of $Y_{b'}$ with edge matrix $\mathcal{S}^{b'b'}$

$$\mathcal{A}_{b'b'.a'} = \begin{array}{r|ccccc} & 1 & 3 & 4 & 5 & 10 \\ \hline 1 & 1 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 0 \\ 5 & 0 & 0 & 0 & 1 & 1 \\ 10 & 0 & 0 & 0 & 0 & 1 \end{array}, \qquad \mathrm{clos}(\mathcal{D}_{b'b'-a'}) = \begin{array}{r|ccccc} & 1 & 3 & 4 & 5 & 10 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 \\ 3 & 0 & 1 & 1 & 1 & 0 \\ 4 & 0 & 1 & 1 & 1 & 0 \\ 5 & 0 & 1 & 1 & 1 & 0 \\ 10 & 0 & 0 & 0 & 0 & 1 \end{array}.$$

A stepwise path derivation of the graph with edge matrix $\mathcal{S}^{b'b'}$ is as follows. The two source-generated edges of the covariance graph with edge matrix $\mathcal{D}_{b'b'-a'}$ arise from the induced subgraphs $3 \leftarrow 8 \rightarrow 4$ and $4 \leftarrow 6 \rightarrow 5$ in Fig. 6(b). After the path from node 3 to 5 via node 4 in this graph is closed by a (3,5)-edge the edges of the concentration graph result, which has edge matrix $\mathrm{clos}(\mathcal{D}_{b'b'-a'})$. This concentration graph is combined with the subgraph that is induced by $b'$ in Fig. 6(b), which has edge matrix $\mathcal{A}_{b'b'.a'}$, to give the graph of Fig. 7(a). Then, for the marginal concentration graph with edge matrix $\mathcal{S}^{b'b'}$, which is shown in Fig. 7(b), two additional edges are induced. This follows from theorem 3, case (a). The corresponding edge inducing paths $3$—$5 \leftarrow 10$ and $4$—$5 \leftarrow 10$ connect the non-adjacent node pairs (3,10) and (4,10) in Fig. 7(a).

Thus for the edge matrices that are discussed in this section the equivalence has been established between an additional 1 in the induced edge matrix and an additional edge generated in the corresponding induced graph, as well as between an edge inducing path in particular graphs or combinations of graphs and the calculation of special sums of edge matrix products which define an additional edge in an induced graph. Next we discuss when an induced edge in a system of triangular densities corresponds to a lack of factorization of the densities.

## 7. Factorizations of densities and dependences induced by triangular systems

We now can prove that a structural 0 correlation, $\rho_{ik.C} = 0$, that is induced in a linear triangular system by a given parent graph implies that the corresponding independence statement, $i \perp\!\!\!\perp k|C$, holds for all possible distributions generated over the same parent graph. This happens because every structural 0 correlation is equivalent to the absence of particular types of edge inducing paths. This absence implies that if a triangular system of densities were to be generated over the same graph then the corresponding factorization of densities is preserved.

The connection of these results with earlier work is in outline as follows. That work focused on properties deriving from the notion of conditional independence and the resulting factorization of densities. For Gaussian distributions the results can be expressed in terms of correlations and partial correlations. Smith (1989) (especially example 3.1) noted that in the proofs of implications of independences only three special properties of conditional independence are usually involved and these properties hold for vanishing correlations and partial correlations and thus apply to linear least squares regression systems. The approach and proof that are used in the present paper are different.

We proceed to show how a non-zero correlation $\rho_{ik.C} \neq 0$ generated via an edge inducing path in a linear triangular system will also lead to a dependence in triangular systems of densities in which some families of distributions are characterized by the technical condition of completeness. Completeness of a family of distributions for a random variable $W$ means that any function of $W$ with zero mean for all distributions in the family is identically 0 (see, for example, Kotz *et al.* (1982)). For instance a regular exponential family is complete.

The essence of the argument is that marginal densities $f_{ik}$ of such families of distributions do not factorize if obtained as $\int f_{i|j} f_{j|k} f_k \,\mathrm{d}F_j$ or $\int f_{i|l} f_{k|l} \,\mathrm{d}F_l$, i.e. by integrating over a common transition or over a common source node respectively. Further the conditional densities $f_{ik|h}$ of such families do not factorize if obtained as

$$f_{h|ik} f_i f_k \bigg/ \int\!\!\int f_{h|ik} \, \mathrm{d}F_i \, \mathrm{d}F_k.$$

The argument derives from a convex combination of separable functions being not itself separable except in degenerate cases. A simple combination of this type is $a_1(x)\, b_1(y) + a_2(x)\, b_2(y)$, where each of the two components is a separable function of variables $x$ and $y$. It is itself of the form $a(x)\, b(y)$ only exceptionally. The following discussion focuses on a set of conditions which excludes such exceptions for densities and which therefore justifies the conclusion that a new dependence is indeed induced.

To derive the results we use the term integration and density for both continuous and discrete variables. We also assume that all random variables have densities such that the calculation of conditional and marginal densities leads to proper distributions. We recall from equations (12) and (16) that in a triangular system of linear equations (13), written as $AY = \varepsilon$, a zero element in position $(i, k)$ of $A$ means $\rho_{ik.\mathrm{par}_i} = 0$, for $\mathrm{par}_i \subseteq \{i+1, \ldots, d_N\}$, and that this is equivalent to a missing $ik$-edge in its parent graph. By contrast a missing $ik$-edge in the parent graph of a triangular system of densities (1), written as $f_N = \Pi_i f_{i|\mathrm{par}_i}$, means that $i \perp\!\!\!\perp k | \mathrm{par}_i$ holds for all distributions that are generated over this parent graph.

For some simple motivating cases we note from the recursion relation for regression coefficients that for three variables $U$, $V$ and $W$ of a linear triangular system and an arbitrary conditioning set $C$ the conditions under which the regression coefficient $\beta_{uv.wC}$ is collapsible over $W$ are that

$$\beta_{uv.wC} = \beta_{uv.C} \qquad \text{if and only if } \rho_{uw.vC} = 0 \text{ or } \rho_{wv.C} = 0. \tag{35}$$

As one consequence, the correlation coefficient that is induced for the end points of an edge inducing path is proportional to the product of the correlation coefficients that are associated with each edge along the path.

We note next that the collapsibility conditions for regression coefficients in condition (35) are $U \perp\!\!\!\perp W | V, C$ or $W \perp\!\!\!\perp V | C$ for a Gaussian distribution and that these coincide with the sufficient conditions for discrete distributions under which constant relative risks for a binary response $U$ with respect to $V$ given $W$ are collapsible over $W$ (Wermuth (1987), proposition 4). Then we note that one of these two types of independence is satisfied at each of the successive V-configurations defining an edge inducing path. In the subsequent arguments concerning distributions we assume a fixed conditioning set $C$ which will not be shown explicitly.

As two examples of joint distributions generated over a four-node path given $C$ we now take

$$U \rightarrow W \leftarrow V \leftarrow X,$$
$$U \rightarrow W \leftarrow V \rightarrow X.$$

Both paths are defined by two successive V-configurations, by those for $(U, W, V)$ and for $(W, V, X)$. In both cases the subgraph that is induced by nodes $U$, $W$ and $V$ is a sink-oriented V-configuration. The V-configuration that is attached to nodes $W$, $V$ and $X$ is transition oriented in the path on the top and source oriented in the path on the bottom. The two joint densities which factorize from equation (1) are then

$$f_{W|UV}\, f_{V|X}\, f_U\, f_X,$$
$$f_{W|UV}\, f_{X|V}\, f_U\, f_V.$$

The paths imply the independence $U \perp\!\!\!\perp X$ as a logical consequence of both generating processes. This may in such simple cases be derived directly from the given density factorizations or from one of the separation criteria formulated in terms of paths. Equivalently, it follows from lemma 4 that in both cases the parent graph induces no edge for pair $(U, X)$ in the overall covariance graph. From theorem 3, case (b), it follows that there is, however, an edge induced for pair $(U, X)$ in the covariance graph of $U$, $V$ and $X$ given $W$. Thus, $U \perp\!\!\!\perp X|W$ is not implied by the parent graphs. For a joint Gaussian distribution the induced partial correlation coefficient $\rho_{ux.w}$ is a positive multiple of $-\rho_{uw}\rho_{wv}\rho_{vx}$. But, there are special types of distribution generated over each of the above two paths which nevertheless satisfy $U \perp\!\!\!\perp X|W$. We now study why this will or will not happen.

*Lemma 5* (conditional and marginal independence). Let three random variables $U$, $V$ and $W$ have density $f_{UVW}$ and let both $U$ and $V$ be marginally dependent on $W$. Then $U \perp\!\!\!\perp V|W$ and $U \perp\!\!\!\perp V$ imply that both $W|U$ and $W|V$ belong to an incomplete family of distributions.

*Proof.* Lemma 5 is proved by using and extending a result (Darroch, 1962) for discrete distributions to general types of distribution and relating it to completeness. In the discrete case, with $p_{uvw}$ denoting the joint probability $\Pr(U=u, V=v, W=w)$ for $u=1, \ldots, I$, $v=1, \ldots, J$ and $w=1, \ldots, K$ and with $p_{.vw}=\Sigma_u p_{uvw}$, and $p_{..w}=\Sigma_v p_{.vw}$ denoting marginal probabilities, Darroch showed that $U \perp\!\!\!\perp V|W$ and $U \perp\!\!\!\perp V$ both hold if and only if the probabilities are from what he called a perfect contingency table. This implies in particular that the bivariate marginal distributions of $U$ and $W$, and $V$ and $W$ are restricted by

$$\sum_w p_{u.w} p_{.vw}/p_{..w} = p_{u..} p_{.v.}.$$

Birch (1963) gave an example for a $2 \times 2 \times 3$ table and Studený (personal communication) an example of a family. This condition generalizes as follows. Under the assumptions of lemma 5, the independences $U \perp\!\!\!\perp V|W$ and $U \perp\!\!\!\perp V$ imply

$$\begin{aligned} f_{UVW} &= f_{UW} f_{VW}/f_W, \\ \int f_{UVW}\, \mathrm{d}_W &= f_U f_V, \end{aligned} \tag{36}$$

where for instance $f_{VW}=\int f_{UVW}\,\mathrm{d}_U$ and $f_W=\int f_{VW}\,\mathrm{d}_V$ denote marginal densities that are derived from $f_{UVW}$. Now condition (36) implies that the distributions of both $W|U$ and of $W|V$ are members of an incomplete family of distributions. For it follows from condition (36) after dividing by $f_V$ that

$$\int (f_{U|W} - f_U)\, f_{W|V}\,\mathrm{d}_W = \int \frac{f_{UW} f_{VW}}{f_W f_V}\,\mathrm{d}_W - f_U \int f_{W|V}\,\mathrm{d}_W = 0. \tag{37}$$

But, if $U$ and $W$ are dependent, then $f_{U|W} \neq f_U$ so the left-hand side specifies a non-zero function of $W$ with zero expectation under the conditional distribution of $W$ given $V=v$ for all $v$. Therefore $W|V$ belongs to an incomplete family of distributions. The same type of argument applies to $W|U$ after dividing expression (36) by $f_U$. □

Darroch's restrictions are never satisfied for a distribution of three binary variables and never for a Gaussian distribution, i.e. simultaneous marginal and conditional independence for any pair, say for $U$ and $V$, can only hold for these distributions if a stronger independence statement is satisfied, i.e. if $U \perp\!\!\!\perp V, W$ or $V \perp\!\!\!\perp U, W$.

*Corollary 3* (dependence induced after changing the conditioning set). Let three random variables $U$, $V$ and $W$ have density $f_{UVW}$, let both $U$ and $V$ be marginally dependent on $W$ and let $W|U$ or $W|V$ be a member of a complete family of distributions. Then

(a) $U \perp\!\!\!\perp V|W$ implies that $U$ is marginally dependent on $V$ and
(b) $U \perp\!\!\!\perp V$ implies that $U$ is conditionally dependent on $V$ given $W$.

*Proof.* Re-expression of lemma 5 with either $W|U$ or $W|V$ being a member of a complete family of distributions gives the implications of cases (a) and (b). □

*Lemma 6* (trivariate association inducing families of distributions). Let a family of trivariate distributions for $U$, $V$ and $W$ be such that both families of the conditional distribution of $W|V$ and of $W|U$ are complete and such that one independence statement is given to hold for pair $(U, V)$. Then this family of distributions is association inducing for pair $(U, V)$, i.e.

(a) $U \perp\!\!\!\perp V|W$ implies that $U$ is marginally dependent on $V$ and
(b) $U \perp\!\!\!\perp V$ implies that $U$ is conditionally dependent on $V$ given $W$.

*Proof.* The assumption of lemma 6 excludes families of distributions for which a single independence statement for $(U, V)$ cannot hold. It excludes further families in which the two bivariate margins of $U$ and $W$, and $V$ and $W$ are constrained by Darroch's restriction. Finally it excludes bivariate margins, of $U$ and $W$ or of $V$ and $W$, that are constrained by an additional independence. If for instance $V$ were marginally independent of $W$, so that $f_{W|V} = f_W$, then equation (37) would be trivially satisfied for $f_{U|W} \neq f_U$, i.e. the distribution of $W|V$ would be a member of an incomplete family. Thus, by the assumed completeness both $U$ and $V$ are marginally dependent on $W$ and the result follows from corollary 3. □

More specifically lemma 6 excludes the family of partially dichotomized Gaussian distributions, i.e. those which are obtained from a joint Gaussian distribution for $U$, $V$ and $W'$ by dichotomizing $W'$ to give $W$, as not sufficiently rich to be association inducing in the case $U \perp\!\!\!\perp V|W$. The reason is that this independence can hold in such a partially dichotomized Gaussian distribution only if either $U \perp\!\!\!\perp W$ or $V \perp\!\!\!\perp W$, in addition to $U \perp\!\!\!\perp V|W$ (see Cox and Wermuth (1992, 1999)). Some trivariate families included as being association inducing are those in which the complete families of $W|U$ and $W|V$ are conditional Gaussian distributions or conditional Gaussian regressions (Lauritzen and Wermuth, 1989). And, these include arbitrary discrete distributions and Gaussian distributions as special cases.

Lemma 6 applied to families of distributions generated over the three types of V-configurations in a parent graph gives a general condition under which a dependence is induced by marginalizing over a transition or a common source node and under which a dependence is induced by conditioning on a common sink node. Put differently, it gives a sufficient condition under which an edge inducing V-configuration is also association inducing, irrespective of the type of V-configuration.

*Theorem 4* (relations induced by triangular systems). For a given parent graph with edge matrix $\mathcal{A} = \mathrm{Ed}[A]$ let $C$ be an arbitrary conditioning set for a pair of nodes $(i, k)$.

(a) Suppose that with this parent graph $\rho_{ik.C} = 0$ is implied for every linear triangular system. Then $i \perp\!\!\!\perp k|C$ is implied for every triangular system of densities that are generated over the same parent graph.
(b) Suppose that with this parent graph a non-zero $\rho_{ik.C}$ can be generated for some linear triangular systems. Then with this parent graph a dependence of $i$ and $k$ given $C$ can also be generated for some members of distributions of abitrary form, provided that all successive trivariate families along at least one edge inducing path for $(i, k)$ given $C$ are association inducing.

*Proof.* For case (a) let $\bar{C}$ denote the nodes outside $C$. No $ik$-edge is induced in the conditional covariance graph of nodes $\bar{C}$ given $C$ because, if such an edge were induced, a non-zero correlation $\rho_{ik.C}$ could be generated. In more detail it follows from theorem 3, case (b), that either

(i) $k$ would be a $\bar{C}$-line ancestor of $i$, or
(ii) $i$ and $k$ would have a common $\bar{C}$-line ancestor, or
(iii) a $\bar{C}$-line ancestor of one of the nodes $i$ or $k$ would have become connected by a sink-generated edge from $\mathrm{clos}(\mathcal{D}^{\bar{C}\bar{C}+C})$ to the other node or
(iv) a $\bar{C}$-line ancestor of $i$ and a $\bar{C}$-line ancestor of $k$ would have become connected by such an edge.

It then follows from the lack of such paths and from the definition of a missing edge in the conditional covariance graph given $C$ for triangular systems of densities that $i \perp\!\!\!\perp k|C$ is implied for any system of densities that are generated over the given parent graph. Put differently, a factorization given in the density (i) for pair $(i, k)$ is retained if there is no edge inducing path for $(i, k)$ given $C$.

For case (b) and a single edge-inducing path for $(i, k)$ given $C$ the claim follows with repeated application of lemma 6, where for each node along the path the marginal family of distributions is also complete. When there are several edge inducing paths the same argument applies to at least one of these paths. □

In the case of several edge inducing paths an independence instead of a dependence may occur for a particular member of the generated family owing to the cancellation of the contributions to the dependence by different paths. Such situations have been called parametric cancellation (Wermuth and Cox, 1998a) or lack of faithfulness of the graph (Spirtes *et al.*, 1993). For a detailed discussion for Gaussian distributions, see Wermuth and Cox (1998a). For related results on the existence of distributions in which an edge inducing path implies dependence, see Geiger and Pearl (1990) and Meek (1995).

In the present paper we provide matrix tools to decide for any chain graph model that is derived from a given triangular system whether an independence statement or a zero correlation is due to parametric cancellation or whether it is a logical consequence of the generating process.

In summary, there are three approaches to the study of both independences and dependences arising from a triangular system: using edge matrices, using paths in graphs or using factorization properties of densities. We have now established the type of intimate relations between the three approaches and can turn to induced chain graph models.

## 8. Induced chain graph models

### 8.1. Parameter matrices induced by linear triangular systems

The results (30)–(32) for three orthogonalized systems can be directly applied to induced linear chain graph models after noting from equation (2) that each chain component $g$ of a chain graph defines an ordered partioning of the node set into three components with $N = (l, g, r)$, where $l = \{1, \ldots, g-1\}$ is the set of nodes in the future of $g$ which we draw to the left of $g$ and where $r = \{g+1, \ldots, d_{\mathrm{CC}}\}$ is the set of nodes in the past of $g$ which we draw to the right of $g$. The following matrix expressions use different splits into two components with $N = (l, R)$, $R = \{g, r\}$, and $N = (L, r)$, $L = \{l, g\}$.

*Theorem 5* (induced parameters of linear regression chains). The induced parameters for component $g$ are of the general form that is given with theorem 2. In detail:

(a) for a linear multivariate regresssion chain they are

$$\Sigma_{gg.r} = [A_{LL}^{-1}(\Delta^{LL+r})^{-1}A_{LL}^{-\mathrm{T}}]_{g,g},$$
$$\Pi_{g|r} = [-A_{LL}^{-1}\Delta_{LL}\theta_{r|L}^{\mathrm{T}}(\Delta_{rr-L})^{-1}A_{rr.L} + \phi_{L|r}]_{g,r}, \tag{38}$$

where $(\Delta^{LL+r})$ is diagonal and equal to the inverse of $\Delta_{LL}$ if $r$ is the empty set;

(b) for a linear blocked concentration chain they are

$$\Sigma^{gg.l} = \Sigma_{gg.r}^{-1} = [A_{RR.l}^{\mathrm{T}}(\Delta_{RR-l})^{-1}A_{RR.l}]_{g,g},$$
$$\Sigma^{gr.l} = [A_{RR.l}^{\mathrm{T}}(\Delta_{RR-l})^{-1}A_{RR.l}]_{g,r}; \tag{39}$$

(c) for a linear concentration regression chain they are $\Sigma^{gg.l}$ and $\Pi_{g|r}$.

*Proof.* The results follow from theorem 3 and the parameter matrices in linear regression chains (see for example Wermuth and Cox (2001)).

### 8.2. *Edge matrices of chain graphs induced by triangular systems of densities*

With the same notation for two different splits of $N$ as used in Section 8.1 and $\mathcal{S}_{gg|r} = \mathrm{Ed}[\Sigma_{gg.r}]$, $\mathcal{P}_{g|r} = \mathrm{Ed}[\Pi_{g|r}]$, $\mathcal{S}^{gg|r} = \mathrm{Ed}[\Sigma^{gg.l}]$ and $\mathcal{C}_{g|r} = \mathrm{Ed}[\Sigma^{gr.l}]$ we can now turn to chain graphs that are induced by triangular systems of densities (1).

*Theorem 6* (edge matrices of induced chain graphs). Edge matrices for component $g$ are of the general form given by theorem 3. In detail:

(a) for a multivariate regresssion chain graph, defined with expressions (2)–(4), they are

$$\mathcal{S}_{gg|r} = \mathrm{In}[\mathcal{A}^{LL}\,\mathrm{clos}(\mathcal{D}^{LL+r})(\mathcal{A}^{LL})^{\mathrm{T}}]_{g,g},$$
$$\mathcal{P}_{g|r} = \mathrm{In}[\mathcal{A}^{LL}\mathcal{T}_{rL}^{\mathrm{T}}\,\mathrm{clos}(\mathcal{D}_{rr-L})\mathcal{A}_{rr.L} + \mathcal{F}_{Lr}]_{g,r}; \tag{40}$$

(b) for a blocked concentration chain graph, defined with expressions (2), (5) and (6), they are

$$\mathcal{S}^{gg|r} = \mathrm{In}[\mathcal{A}_{RR.l}^{\mathrm{T}}\,\mathrm{clos}(\mathcal{D}_{RR-l})\mathcal{A}_{RR.l}]_{g,g},$$
$$\mathcal{C}_{g|r} = \mathrm{In}[\mathcal{A}_{RR.l}^{\mathrm{T}}\,\mathrm{clos}(\mathcal{D}_{RR-l})\mathcal{A}_{RR.l}]_{g,r}; \tag{41}$$

(c) for a concentration regression chain they are $\mathcal{S}^{gg|r}$ and $\mathcal{P}_{g|r}$.

*Proof.* The results are a direct consequence of theorems 4 and 5. □

In summary, 0s in induced edge matrices in equations (40) and (41) identify structural 0s in induced parameter matrices (38) and (39) of linear systems and missing edges in chain graphs, as well as independences in chain graph models implied by triangular systems of densities (1).

## 9. Using the results

### 9.1. *Joint response graphs induced after explicitly marginalizing and conditioning*

We now illustrate induced chain graphs that are obtained after conditioning on a subset $C$ of the variables and after marginalizing over another subset $M$ so that only the remaining variables are of interest. The factorization (2) of the density in a chain graph model implies that marginalizing over the first component and conditioning on the last will leave the densities of the remaining variables unchanged.

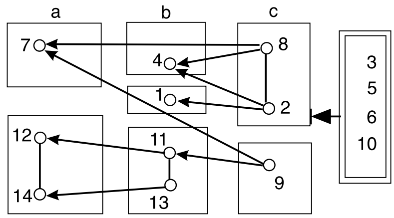


**Fig. 8.** Chain graph of blocked concentrations induced by the parent graph of Fig. 3 for $N=(a,b,c,C)$ with $a=\{7,12,14\}$, $b=\{1,4,11,13\}$, $c=\{2,8,9\}$ and $C=\{3,5,6,10\}$, shown after having conditioned on $C$

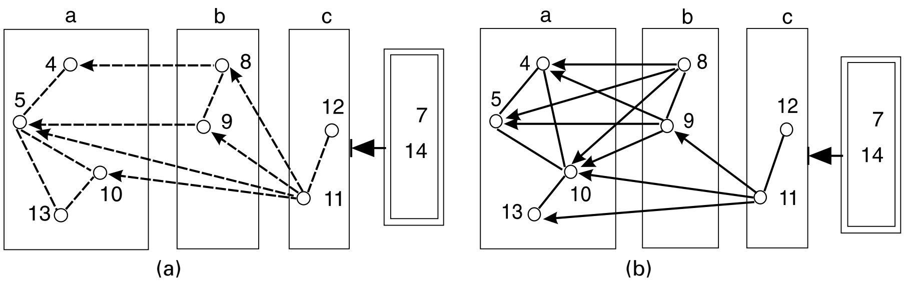


**Fig. 9.** Two chain graphs induced by the parent graph of Fig. 3 for $N=(M,a,b,c,C)$ with $M=\{1,2,3,6\}$, $a=\{4,5,10,13\}$, $b=\{8,9\}$, $c=\{11,12\}$ and $C=\{7,14\}$, shown after having marginalized over $M$ and conditioned on $C$: (a) for multivariate regressions and (b) for blocked concentrations

For the blocked concentration graph that is shown in Fig. 8 and induced by the parent graph in Fig. 3 the overall node set $N$ containing 14 nodes is ordered with the selected chain components as $N=(a,b,c,C)$. The nodes within each of the four components are as shown in Fig. 8. The set $C$ denotes an overall conditioning set. The factorization of the density due to the chosen chain is given by equation (2), the meaning of the edges of the chain graph type is defined by expressions (5) and (6), and the edge matrix and hence the edges that are present and absent in Fig. 8 are given by case (b) of theorem 6.

The stacked boxes in Fig. 8 indicate mutual independence of components of the joint responses within a given chain component conditionally given $C$ and the variables in chain components $g+1,\ldots,d_{\mathrm{CC}}$. For instance, for chain component $b$ of Fig. 8 we have the independence $(4 \perp\!\!\!\perp 1 \perp\!\!\!\perp 11,13)|\{2,8,9,C\}$. By theorem 6 three further arrows are added to both the induced multivariate regression and the concentration regression chain: for (13,9), (12,13) and (14,11). Thus, in this example the three types of induced chain graph are similar.

By contrast the two induced chain graphs in Fig. 9, both for the same partitioning $N=(M,a,b,c,C)$, differ much in the edges that are present. For instance, node 10 has three neighbours in the multivariate regression chain in Fig. 9(a), but six in the blocked concentration chain in Fig. 9(b). In this example the concentration chain is considerably more complex than the multivariate regression chain.

### 9.2. *A matrix criterion for separation in triangular systems of densities*

The equivalence of the following matrix criterion for separation to the path criteria that have been given previously in the literature has been proved elsewhere.

*Corollary 4* (separation in triangular systems). For triangular systems of densities (1) the following statements for three disjoint node subsets $\alpha$, $\beta$ and $C$ of $N$ are equivalent:

(a) the independence $\alpha \perp\!\!\!\perp \beta | C$ is implied by the parent graph;
(b) no edge connects the subgraph of nodes $\alpha$ and the subgraph of nodes $\beta$ in the induced covariance graph for $\alpha$ and $\beta$ given $C$;
(c) the edge matrix of the induced covariance graph for $(\alpha, \beta)$ given $C$ is block diagonal.

*Proof.* The results follow from theorem 6. □

Note that the edge matrix for the induced conditional covariance graph can be computed with $S = \{a, b\}$ and $\bar{C} = N \backslash C$ from theorem 3 as $\mathcal{S}_{SS|C} = \mathrm{In}[\mathcal{A}^{\bar{C}\bar{C}} \operatorname{clos}(\mathcal{D}^{\bar{C}\bar{C}+C})(\mathcal{A}^{\bar{C}\bar{C}})^{\mathrm{T}}]_{S,S}$. After rearranging the nodes of $\alpha$ to correspond to rows $1$–$d_\alpha$ of this edge matrix and those of $\beta$ to the remaining rows, block diagonality in $\alpha$ and $\beta$ is seen to be equivalent to the required factorization of the joint conditional density as $f_{S|C} = f_{\alpha|C} f_{\beta|C}$. Note that from equation (28) the corresponding covariance matrix that is induced for a given linear triangular system (13) is $\Sigma_{SS.C} = [A^{-1}_{\bar{C}\bar{C}} (\Delta^{\bar{C}\bar{C}+C})^{-1} A^{-T}_{\bar{C}\bar{C}}]_{S,S}$, from which the direction and strength of induced linear associations can be judged in terms of corresponding induced partial correlations.

The edge matrix of the chain graph that is induced by a triangular system, be it a covariance graph as in corollary 4 or a more general graph as in theorem 6, indicates which of the independences in the chain graph model are logical consequences of the generating process. All remaining independences are then due only to special constellations in a given generating system. Indicator matrices of particular sums of products of edge matrices define the induced chain graphs that were discussed here. With each such sum of products the number of edge inducing paths of a particular kind is computed for the variable pairs contained in the induced graph. This knowledge is relevant for judging the absence or possible presence of confounding effects: with only one edge inducing path no confounding of an induced association is possible.

In addition, for linear triangular systems the direction and strength of the relevant induced correlations are given in theorem 5 and are presented in the same matrix form as the edge matrix. This may be relevant also for linear approximations to triangular systems of densities, which are of interest when interactive effects are absent and non-linearities are weak.

## Acknowledgements

We are grateful to the German Research Society, the Australian National University and the Radcliffe Institute for Advanced Study at Harvard University for supporting this work, to Arthur Dempster for searching questions and stimulating discussions of earlier versions, to referees and to Giovanni Marchetti and Thomas Richardson for most helpful comments on how to improve the presentation of the results. Algorithms for computing edge matrices of induced chain graphs have been programmed as macros in MATLAB, distributed by Mathworks Inc., and Johannes Martin has provided Java-based Internet access to them at the Web address `http://psystat.sowi.uni-mainz.de`.

## References

Birch, M. W. (1963) Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc.* B, **25**, 220–233.
Cochran, W. G. (1938) The omission or addition of an independent variate in multiple linear regression. *J. R. Statist. Soc.*, suppl., **5**, 171–176.
Cox, D. R. and Wermuth, N. (1992) Response models for mixed binary and quantitative variables. *Biometrika*, **79**, 441–461.
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*. London: Chapman and Hall.

Cox, D. R. and Wermuth, N. (1999) Likelihood factorizations for mixed discrete and continuous variables. *Scand. J. Statist.*, **25**, 209–220.
Cramér, H. (1946) *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
Darroch, J. N. (1962) Interactions in multi-factor contingency tables. *J. R. Statist. Soc.* B, **24**, 251–263.
Dempster, A. P. (1969) *Elements of Continuous Multivariate Analysis*. Reading: Addison-Wesley.
Edwards, D. (2000) *Introduction to Graphical Modelling*, 2nd edn. New York: Springer.
Frydenberg, M. (1990) The chain graph Markov property. *Scand. J. Statist.*, **17**, 333–353.
Geiger, D. and Pearl, J. (1990) On the logic of causal models. In *Uncertainty in Artificial Intelligence IV* (eds R. D. Schachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer), pp. 136–147. Amsterdam: North-Holland.
Green, P. J., Hjort, N. L. and Richardson, S. (2003) (eds) *Models for Highly Structured Stochastic Systems*. Oxford: Oxford University Press.
Koster, J. T. A. (2002) Marginalizing and conditioning in graphical models. *Bernoulli*, **8**, 817–840.
Kotz, S., Johnson, N. L. and Read, C. (1982) Completeness. In *Encyclopedia of Statistical Sciences* (eds S. Kotz, N. L. Johnson and C. Read), vol. 2, p. 82. New York: Wiley.
Lauritzen, S. L. (1996) *Graphical Models*. Oxford: Oxford University Press.
Lauritzen, S. L., Dawid, A. P., Larsen, B. and Leimer, H.-G. (1990) Independence properties of directed Markov fields. *Networks*, **20**, 491–505.
Lauritzen, S. L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31–54.
Levitz, M., Perlman, M. D. and Madigan, D. (2001) Separation and completeness properties for AMP chain graph Markov models. *Ann. Statist.*, **29**, 1751–1784.
Meek, C. (1995) Strong completeness and faithfulness in Bayesian networks. In *Proc. 11th A. Conf. Uncertainty in Artificial Intelligence*, pp. 411–418. San Francisco: Morgan Kaufmann.
Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*. San Mateo: Morgan Kaufmann.
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 Markov graphical models. *Ann. Statist.*, **30**, 962–1030.
Searle, S. R. (1966) *Matrix Algebra for the Biological Sciences*. New York: Wiley.
Smith, J. Q. (1989) Influence diagrams for statistical modelling. *Ann. Statist.*, **17**, 654–672.
Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction and Search*. New York: Springer.
Studený, M. and Bouckaert, R. (1998) On chain graph models for description of conditional independence structures. *Ann. Statist.*, **26**, 1434–1495.
Wermuth, N. (1987) Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. R. Statist. Soc.* B, **49**, 353–364.
Wermuth, N. (1992) On block-recursive linear regression equations (with discussion). *Rev. Bras. Probil. Estatist.*, **6**, 1–56.
Wermuth, N. (2003) Analysing social science data with graphical Markov models. In *Highly Structured Stochastic Systems* (eds P. Green, N. Hjort and S. Richardson), pp. 47–52. Oxford: Oxford University Press.
Wermuth, N. and Cox, D. R. (1998a) On association models defined over independence graphs. *Bernoulli*, **4**, 477–495.
Wermuth, N. and Cox, D. R. (1998b) Statistical dependence and independence. In *Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton), pp. 4260–4267. Chichester: Wiley.
Wermuth, N. and Cox, D. R. (2001) Graphical models: overview. In *International Encyclopedia of the Social and Behavioral Sciences* (eds P. B. Baltes and N. J. Smelser), vol. 9, pp. 6379–6386. Amsterdam: Elsevier.
Wermuth, N. and Lauritzen, S. L. (1990) On substantive research hypotheses, conditional independence graphs and graphical chain models (with discusssion). *J. R. Statist. Soc.* B, **52**, 21–72.
Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.