A sweep algorithm for triangular matrices and its statistical applications

Wermuth, Nanny; Cox, D.R.

ZUMA-Arbeitsbericht 00/04

A sweep algorithm for triangular matrices and its statistical applications

Nanny Wermuth und D.R. Cox

Juni 2000

ISSN 1437-4110

ZUMA
Quadrat B2,1
Postfach 12 21 55
68072 Mannheim
Telefon: 0621-1246-155
Telefax: 0621-1246-100
E-Mail: wermuth@zuma-mannheim.de

A sweep operator for triangular matrices and its statistical applications *

Nanny Wermuth
Center of Survey Research (ZUMA), Postfach 122155, Mannheim, Germany. e-mail: wermuth@zuma-mannheim.de
D.R.Cox
Department of Statistics and Nuffield College, Oxford OX1 1NF, UK. e-mail: david.cox@nuf.ox.ac.uk

SUMMARY

A sweep operator is defined for stepwise sequential inversion of triangular matrices and its properties are compared to those of the sweep operator for inverting symmetric matrices. The algorithm is used to study joint distributions generated over a directed acyclic graph. Three main applications are derived. The first is to prove a simple form for the joint distribution resulting after marginalising over and conditioning on arbitrary subsets of variables in such a linear system. The second is to extend the results for linear systems to general distributions by interpreting structural zeros in matrices in terms of missing edges in associated graphs and symbolic matrix transformations as modifications of graphs. The third is to show the equivalence of several criteria for reading off independence statements from directed acyclic graphs.

Keywords: connected dependencies; graphical Markov model; independence graph, joint response model; separation criteria; triangular matrix decomposition; univariate recursive regression

*Research Report, ZUMA 00-04; as submitted to Bernoulli

1. Generating processes and independence graphs

Association models may be defined over an independence graph (Wermuth & Cox, 1998), which has a vertex or node set V of d_V elements, where the nodes correspond to random variables $Y_1, \ldots, Y_{d_V}$. The graph is called a directed, acyclic graph and denoted by G^V_{dag}, if each pair of nodes has at most one edge, each edge is an arrow and starting from any one node it is impossible to return to this node by following only directions in which the arrows point. Figure 1 shows a directed acyclic graph in ten nodes that we shall use to illustrate results.

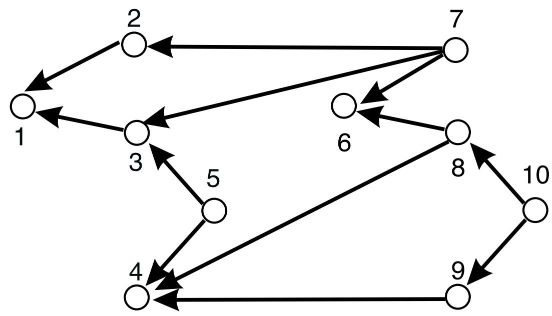


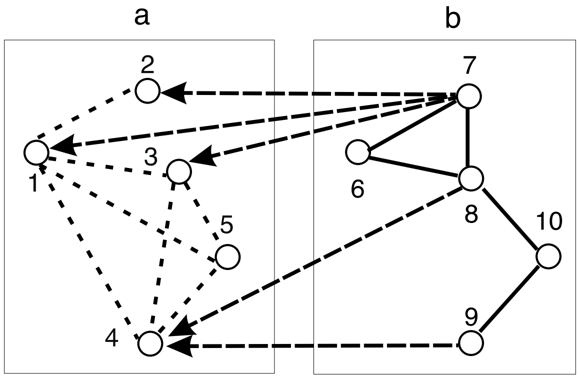
Figure 1: A directed acyclic graph in 10 nodes

Each directed acyclic graph may describe independencies in a univariate recursive generating process for a joint distribution, in which variable Y_1 denotes the most recent response variable; Y_{d_V} denotes the last, purely explanatory variable; all other variables are intermediate since they may play the role of responses to some and the role of explanatory variables to some other variables in the system. A *generating process* starts with the marginal distribution of the purely explanatory variable and generates the conditional distribution of response variable Y_i for $i = 1, \ldots, d_V - 1$, in terms of a subset of $Y_{i+1}, \ldots, Y_{d_V}$. The generating process provides a full ordering of the nodes $(1, 2, \ldots, d_V)$ such that the joint density f_V factorizes accordingly into d_V univariate (conditional) densities as

$$f_{1,\ldots,d_V}(Y_1, \ldots, Y_{d_V}) = f_{d_V}(Y_{d_V})\Pi_{i=1}^{d_V-1} f_i(Y_i \mid Y_{\text{par}(i)} = y_{\text{par}(i)}). \tag{1}$$

Here $Y_{\text{par}(i)}$ is the subset of $\{Y_{i+1}, \ldots, Y_{d_V}\}$ for which arrows point in G^V_{dag} directly to node i. They are the directly explanatory variables; the corresponding nodes are

Figure 2: A joint response chain graph as implied by Figure 1 for Y_a given Y_b with $(a, b) = (1, \ldots, 10)$; a conditional covariance graph for Y_a given Y_b; arrows for each component of Y_a projected on Y_b; a concentration graph for the marginal distribution of Y_b



called the *parent* nodes of i. The nodes in the set $\{i, i+1, \ldots, d_v\}$ are *potential ancestors* of i in a given generating process. They are to be distinguished from the proper ancestors. Node j is an *ancestor* of node i, i.e. a proper one, if a direction-preserving path leads from j to i. A *path* of length n is a succession of n edges connecting nodes $i_0, \ldots, i_n$, irrespective of the orientation of the edges. For instance in Figure 1 nodes $7, 8, 10$ are ancestors of node 6, nodes $7, 8$ are its parents and nodes $7, 8, 9, 10$ are its potential ancestors. Whenever j is an ancestor of i, node i is called a *descendant* of node j.

The factorization (1) of the joint density specifies for each $i < d_V$ the joint density of $Y_i, \ldots, Y_{d_V}$ written in condensed form as

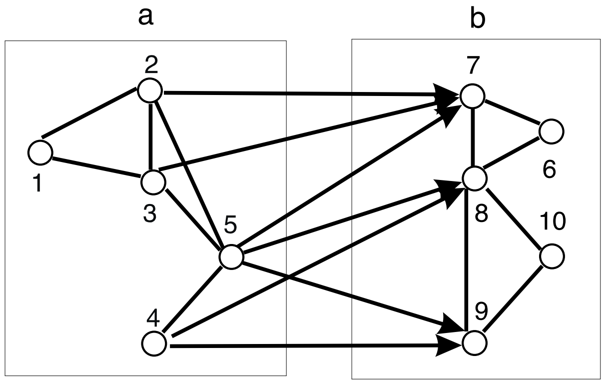
$$f_{i,\ldots,d_V} = f_{i|\mathrm{par}(i)} f_{i+1,\ldots,d_V}$$

so that $f_{i|i+1,\ldots,d_V}$, the conditional density of Y_i given its potential ancestors, depends only on $Y_{\mathrm{par}}(i)$. Therefore the defining independence structure may be written in terms of response variables by referring only to the nodes and edges in G^V_{dag} as

$$\{i \perp\!\!\!\perp \text{ (potential ancestors of } i \text{ excluding parents of } i) \mid \text{ parents of } i\}. \qquad (2)$$

The information on the factorization of a density (1) may, equivalently, be stored in the *edge matrix* of the graph, an indicator matrix of zeros and ones. The edge

Figure 3: A joint response chain graph as implied by Figure 1 for Y_b given Y_a with $(a, b) = (1, \ldots, 10)$; a block-regression graph for Y_b given Y_a; a concentration graph for the marginal distribution of Y_a



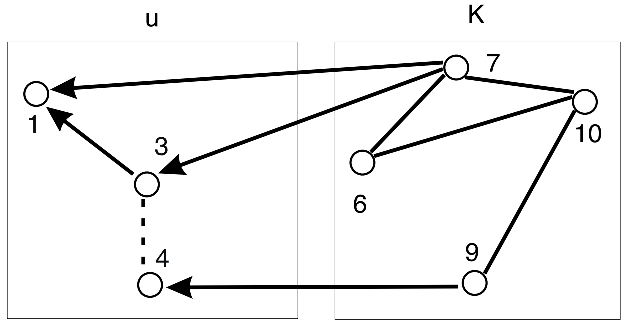
matrix, $\mathcal{A}$ of a directed acyclic graph, G^V_{dag} has ones along the diagonal, zeros below the diagonal, and a nonzero entry for node pair (i, j) if and only if node j is a parent of node i, or, equivalently, if Y_j is directly explanatory for Y_i. Because of this property the directed acyclic graph of a generating process is sometimes called the *parent graph.* Here and in the following we use the conventions that nodes (h, i, j, k) are ordered so that h has the smallest and k the largest numbering of the four nodes and that node 1 in the generating process corresponds to the first row and node d_V to the last row of $\mathcal{A}$.

We shall use further edge matrices of graphs such as the edge matrix of an *ancestor graph.* It has a nonzero (i, j)-entry if and only if node j is an ancestor of node i in the corresponding parent graph.

As we shall see, transformations on edge matrices help us to understand independence structures for joint response models induced by systems of univariate recursive regressions. Figures 2 to 4 present examples of such independence structures resulting from the directed acyclic graph in Figure 1. We shall show how to derive and interpret them, first for linear systems by transforming matrices, then for general systems by transforming graphs, or equivalently, their edge matrices.

In Section 2 we introduce a new sweep operator for triangular matrices to supplement the usual sweep operator for symmetric matrices. Both permit us to perform

Figure 4: A chain graph of connected univariate dependencies as implied by Figure 1 for Y_u, $u \subset a$ conditional given Y_K, $K \subset b$ with $(a, b) = (1, \ldots, 10)$; a directed acyclic graph permitting arrows or dashed lines as edges or both as edge components for the conditional distribution of Y_u given Y_K, a transformed matrix of projecting each component of Y_u on Y_K and a concentration graph for the marginal distribution of Y_K



certain matrix manipulations such as inversion by repeated steps involving three positions one at a time. In Section 3 we study some key properties of Gaussian systems defined via a linear recursive system and in Section 4 we give in matrix form the properties for a new linear system derived by a combination of marginalising over some variables and conditioning on others. In Section 5 the relation between the matrix transformations and paths in associated graphs is developed. This leads to the conclusion that independencies obtained as structural properties of all Gaussian systems generated over a directed acyclic graph apply to arbitrary distributions generated over the same graph, essentially because of the factorization property (1). Section 6 gives results about different but equivalent separation criteria for directed acyclic graphs.

Thus one central theme of the paper concerns the interplay between properties of matrices and of associated graphs, derived from a triangular matrix defining a linear system over a generating directed acyclic graph, G^V_{dag}. Among our conclusions are quantitative results for the direction and strength of dependencies obtained by conditioning on some variables and marginalising over others in such linear systems. Further we show the remarkable result that if a conditional or marginal independency holds for all possible Gaussian distributions generated over the graph G^V_{dag} then it holds for all possible distributions generated over the same graph. For deciding on whether any given conditional independence statement is implied by the generating graph a matrix algorithm is given which involves nothing but replacing in a systematic fashion zeros by ones in the edge matrix of the generating graph.

2. Sweep operators

A sweep-operator had been designed by Beaton (1964) as a tool for inverting symmetric matrices. It has nice properties, since it defines stepwise changes of the matrix which can be readily undone by a corresponding resweep-operator and it is an efficient way to successively orthogonalize a symmetric matrix (Dempster, 1969, Section 4.3). As such it is closely related to Gram-Schmidt orthogonalisation and it gives a Cholesky-factorization of a symmetric matrix. For completeness, definitions are repeated in Appendix 1.

We define here a simpler sweep operator for triangular matrices having unit diagonal elements. With the original sweep operator a wealth of matrix identities related to multivariate least squares regressions can be derived. Similarly, with the sweep operator for triangular matrices we derive joint response models generated from systems of linear univariate recursive regressions. More important is that triangular matrices which are swept on some or on all of their indices can be reinterpreted as modifications of directed acyclic graphs. This means that results obtained by sweeping on linear systems generalize to distributions of arbitrary type provided they are generated over directed acyclic graphs.

Let an upper-triangular $r \times r$ matrix A have ones as diagonal elements and let its element in position (i, j) be denoted by a_{ij}. Then a simple sweeping step on row and column i gives another $r \times r$ upper triangular matrix $\tilde{A}$ in which for $h < i < j$:

$$\tilde{a}_{ij} = -a_{ij}, \quad \tilde{a}_{hj} = a_{hj} - a_{hi}a_{ij}, \tag{3}$$

and all other elements of $\tilde{A}$ coincide with those of A.

In this way possibly none, but at most $r^2/4$, elements of A are modified. We write the simple sweeping step on a triangular matrix as (A swt i) to distinguish it from the original sweep operator for symmetric matrices M, which is abbreviated as $(M \text{ swp } i)$.

The following main properties of the swt-operator result by direct calculations.

(1) Sweeping the matrix A on all rows and columns gives the inverse matrix:

$$B = A^{-1} = (\ldots((A \text{ swt } 1) \text{ swt } 2)\ldots \text{swt } r)$$

(2) the order of sweeping can be interchanged without altering the result:

$$((A \text{ swt } i) \text{ swt } j) = ((A \text{ swt } j) \text{ swt } i)$$

(3) sweeping on i of A is undone by reapplying the swt-operator on i of $\tilde{A}$:

$$A = ((A \text{ swt } i) \text{ swt } i).$$

For instance, for a 5×5 matrix A, the inverse B can be written as

$$B = \begin{pmatrix} 1 & -a_{12} & -a_{13.2} & -a_{14.23} & -a_{15.234} \\ 0 & 1 & -a_{23} & -a_{24.3} & -a_{25.34} \\ 0 & 0 & 1 & -a_{34} & -a_{35.4} \\ 0 & 0 & 0 & 1 & -a_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where e.g. $a_{13.2} = a_{13} - a_{12}a_{23}$, $a_{24.3} = a_{24} - a_{23}a_{34}$, and, for instance $a_{14.23}$ can be computed in a number of different ways depending on the order in which the sweeping is carried out

$$a_{14.23} = a_{14.3} - a_{12}a_{24.3} = a_{14.2} - a_{13.2}a_{34}$$

and it will be a *structural zero* i.e. it will be zero for all possible values of free parameters in A, if and only if all individual terms in the following sum vanish

$$a_{14} + a_{12}a_{24} + a_{13}a_{34} + a_{12}a_{23}a_{34}.$$

The notion of a structural zero is essential in this paper. A structural zero contrasts with a zero that occurs only for special constellations in any given set of parameters.

By direct extension we denote by (A swt a) the matrix A swept on the set of rows and columns a. As we shall see, there is a quite close connection between sweeping over rows and columns a and the operation of marginalising over the distribution of random variables in a corresponding linear system of equations discussed in the next Section.

One matrix identity for the matrices A and $B = A^{-1}$ partitioned into two parts a and b results for instance from (A swt a)=(B swt b). When rows and columns are thereby split into two adjacent components so that $(1, \ldots, r) = (a, b)$, i.e. $a = (1, \ldots, r^*)$, and $b = (r^* + 1, \ldots, r)$, this gives

$$\begin{pmatrix} A_{aa}^{-1} & -A_{aa}^{-1}A_{ab} \\ 0 & A_{bb} \end{pmatrix} = \begin{pmatrix} B_{aa} & B_{ab}B_{bb}^{-1} \\ 0 & B_{bb}^{-1} \end{pmatrix}.$$

3. Some joint response models derived from linear recursive systems

In the special case in which the directed acyclic graph G^V_{dag} corresponds to a linear system in Gaussian variables the joint Gaussian distribution is generated by a set of linear recursive regressions with independent residuals. Such linear systems had been introduced in genetics as models of path analysis (Wright, 1921, 1934) and were generalized in psychometrics to linear structural relation models (Jöreskog, 1981). These systems had also been advocated and studied in econometrics (Wold & Jureen, 1953, pp. 48-53; Wold, 1959). They are treated there as a subclass of simultaneous equation models having some appealing features (Goldberger, 1964, pp. 354-355). In the context of graphical Markov models (Lauritzen, 1996; Edwards, 1995; Cox & Wermuth, 1993; 1996) the relation between structural equation models and cyclic independence graphs has been derived by Spirtes (1995) and Koster (1996, 1999a) and Spirtes et al. (1998).

In the ith regression equation in a linear system generated over G^V_{dag} the parameters are regression coefficients and the residual variance obtained when regressing Y_i on $Y_{\text{par}(i)}$. The regression coefficient of Y_j for $j \in \text{par}(i)$ is written as $\beta_{ij.\text{par}(i)\setminus j}$, the residual variance as $\sigma_{ii.\text{par}(i)}$. We assume without loss of generality that all components of Y have mean zero. In matrix notation the whole system is

$$AY = \varepsilon, \quad \mathrm{cov}(\varepsilon) = \Delta, \tag{4}$$

where A is upper triangular with elements $a_{ij} = -\beta_{ij.\mathrm{par}(i)\setminus j}$ and Δ is diagonal with elements $\delta_{ii} = \sigma_{ii.\mathrm{par}(i)}$.

The covariance matrix Σ of Y_V and the concentration matrix Σ^{-1} of Y_V are obtained from $\mathrm{cov}(Y_V) = B\mathrm{cov}(\varepsilon)B^T$ as

$$\Sigma = B\Delta B^T, \quad \Sigma^{-1} = A^T\Delta^{-1}A, \tag{5}$$

i.e. (A, Δ^{-1}) is the triangular decomposition of the concentration matrix and (B, Δ) is the triangular decomposition of the covariance matrix corresponding to the ordering $(1, \ldots, d_V)$. For variables with a joint Gaussian distribution a structural (i, j)-zero in Σ means $Y_i \perp\!\!\!\perp Y_j$, one in Σ^{-1} means $Y_i \perp\!\!\!\perp Y_j \mid Y_{V\setminus\{i,j\}}$ (see e.g. Cox & Wermuth, 1996, p. 69).

Direct calculations show that the matrix product $B\Delta$ has as elements in the upper-triangular part partial variances along the diagonal and partial covariances elsewhere, while $A\Delta^{-1}$ has as diagonal elements partial precisions and negative values of partial covariances, elsewhere. If the generating process for the distribution of Y_V leads to a *saturated model*, i.e. if A has no structural zeros, then $\mathrm{par}(i) = i + 1, \ldots, d_V$, and with structural zeros the factorization in (1) implies that $\sigma_{ii.i+1,\ldots,d_V} = \sigma_{ii.\mathrm{par}(i)}$.

For example, for Y_V having five components and A being without structural zeros the two matrices are

$$A\Delta^{-1} = \begin{pmatrix} \sigma^{11} & -\sigma_{12.345} & -\sigma_{13.245} & -\sigma_{14.235} & -\sigma_{15.234} \\ 0 & \sigma^{22} & -\sigma_{23.45} & -\sigma_{24.35} & -\sigma_{25.34} \\ 0 & 0 & \sigma^{33} & -\sigma_{34.5} & -\sigma_{35.4} \\ 0 & 0 & 0 & \sigma^{44} & -\sigma_{45} \\ 0 & 0 & 0 & 0 & \sigma^{55} \end{pmatrix},$$

where σ^{ii} denotes the precision in the i-th linear regression equation which is the reciprocal value of the residual variance, $\sigma^{ii} = 1/\sigma_{ii.i+1,\dots,d_V}$;

$$B\Delta = \begin{pmatrix} \sigma_{11.2345} & \sigma_{12.345} & \sigma_{13.45} & \sigma_{14.5} & \sigma_{15} \\ 0 & \sigma_{22.345} & \sigma_{23.45} & \sigma_{24.5} & \sigma_{25} \\ 0 & 0 & \sigma_{33.45} & \sigma_{34.5} & \sigma_{35} \\ 0 & 0 & 0 & \sigma_{44.5} & \sigma_{45} \\ 0 & 0 & 0 & 0 & \sigma_{55} \end{pmatrix}.$$

For joint Gaussian distributions defined by (4) this representation implies that every structural zero in A and in B is equivalent to a specific independence statement, since for Gaussian distributions Y_h is independent of Y_k given Y_C if and only if the conditional covariance vanishes for the pair, i.e. $\sigma_{hk.C} = 0$.

If the generating process is for example characterized by the *subgraph induced by nodes* $1, \dots, 5$ in Figure 1, i.e. it is the graph obtained by keeping just these nodes and its edges, then A and B have structural zeros and free parameters as given by

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & a_{35} \\ 0 & 0 & 0 & 1 & a_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & b_{12} & b_{13} & 0 & b_{15} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & b_{35} \\ 0 & 0 & 0 & 1 & b_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6}$$

For instance, for the two structural zeros in row 1 of A this implies $\sigma_{14.235} = \sigma_{15.234} = 0$, so that $Y_1 \perp\!\!\!\perp (Y_4, Y_5) \mid (Y_2, Y_3)$ and the three structural zeros in column 3 of B mean that $\sigma_{14.5} = \sigma_{24.5} = \sigma_{34.5} = 0$ so that the independence statement $Y_4 \perp\!\!\!\perp (Y_1, Y_2, Y_3) \mid Y_5$ follows.

To derive the edges displayed in Figure 2 for a Gaussian system we equate the result of sweeping Σ on indices b (Σ swp b) to the form resulting from (5) after partitioning into two adjacent blocks with $(1, \dots, d_V) = (a, b)$

$$\begin{pmatrix} \Sigma_{aa.b} & \Pi_{a|b} \\ . & -\Sigma_{bb}^{-1} \end{pmatrix} = \begin{pmatrix} B_{aa}\Delta_{aa}B_{aa}^T & B_{ab}B_{bb}^{-1} \\ . & -A_{bb}^T\Delta_{bb}^{-1}A_{bb} \end{pmatrix}, \tag{7}$$

where $\Pi_{a|b}$ denotes the matrix of regression coefficients resulting from linear regression of Y_a on Y_b and dots in the left hand lower corner indicate that we have a symmetric matrix. The matrix $\Pi_{a|b}$ can be computed in a number of different ways, some of which may be derived from (Σ swp b) $=$ ($-\Sigma^{-1}$ rswp a), from (B swt

$b) = (A \text{ swt } a)$, and $\Pi_{a|b} = \Sigma_{ab}\Sigma_{bb}^{-1} = (B_{ab}\Delta_{bb}B_{bb}^T)(B_{bb}\Delta_{bb}B_{bb}^T)^{-1}$ to give

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

The structural zeros and the free parameters in $\Pi_{a|b}$ can thus be determined via the matrix product of $A_{aa}^{-1} = B_{aa}$, (containing the information on ancestors of a within a) and A_{ab} (containing the information on the parents of a in b). In Figure 1 for $a = 1, \ldots, 5$ and b all remaining indices this gives $B_{aa}A_{ab} = \Pi_{a|b}$ as

$$\begin{pmatrix} 1 & b_{12} & b_{13} & 0 & b_{15} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & b_{35} \\ 0 & 0 & 0 & 1 & b_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a_{27} & 0 & 0 & 0 \\ 0 & a_{37} & 0 & 0 & 0 \\ 0 & 0 & a_{48} & a_{49} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \pi_{17} & 0 & 0 & 0 \\ 0 & \pi_{27} & 0 & 0 & 0 \\ 0 & \pi_{37} & 0 & 0 & 0 \\ 0 & 0 & \pi_{48} & \pi_{49} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (8)$$

By replacing nonzero entries in a matrix of regression coefficients by ones, an *indicator matrix of structural zeros* is obtained which is here the edge matrix between blocks in a graph representing multivariate regression of Y_a on Y_b.

The linear equations reflected in the conditional distribution of Y_a given $Y_b = y_b$ in Figure 2 are sometimes called the *reduced form equations* of the recursive system (1). The linear equations for the joint distribution of Y_V are then

$$Y_a - \Pi_{a|b}y_b = B_{aa}\varepsilon_a, \quad Y_b = B_{bb}\varepsilon_b \quad (9)$$

The graph of the marginal distribution of Y_b in Figure 2 represents a concentration graph model, which for Gaussian variables was introduced under the name of covariance selection by Dempster (1972).

The linear equations corresponding to the conditional distribution of Y_b given Y_a may also be expressed in terms of the parameters of the recursive system (1) and a corresponding reduced form. However, this is less direct, since the order of dependencies of the generating system is thereby reversed.

With the matrix in (7) reswept on b and swept further on a the covariance matrix and the concentration matrix of all variables can be expressed in partitioned form as

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ . & \Sigma_{bb} \end{pmatrix} = \begin{pmatrix} B_{aa}\Delta_{aa-b}B_{aa}^T & B_{ab}\Delta_{bb}B_{bb}^T \\ . & B_{bb}\Delta_{bb}B_{bb}^T \end{pmatrix}, \quad (10)$$

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{aa.b}^{-1} & -\Sigma_{aa.b}^{-1}\Pi_{a|b} \\ . & \Sigma_{bb.a}^{-1} \end{pmatrix} = \begin{pmatrix} A_{aa}^T\Delta_{aa}^{-1}A_{aa} & A_{aa}^T\Delta_{aa}^{-1}A_{ab} \\ . & A_{bb}^T\Delta^{bb+a}A_{bb} \end{pmatrix},$$

where $\Delta_{aa-b} = \Delta_{aa} + \Theta_{a|b}\Delta_{bb}\Theta_{a|b}^T$ is the covariance matrix of residuals obtained for a of system (4) after having marginalised over b, i.e. of $\varepsilon_{a-b} = \varepsilon_a - \Theta_{a|b}\varepsilon_b$ and where $\Delta^{bb+a} = \Delta_{bb}^{-1} + \Theta_{a|b}^T\Delta_{aa}^{-1}\Theta_{a|b}$ is the concentration matrix of ε_b in the joint distribution of ε_b and ε_{a-b}. Furthermore, $-\Theta_{a|b} = -A_{aa}A_{bb}^{-1}$ is the matrix of regression coefficients obtained in the linear regression of ε_{a-b} on ε_b, and $\Pi_{b|a} = \Sigma_{ab}^T\Sigma_{aa}^{-1} = -B_{bb}(\Delta_{bb}\Theta_{a|b}^T\Delta_{aa-b}^{-1})A_{aa}$ is the matrix of regression coefficients resulting from linear regression of Y_b on Y_a. Since

$$A_{bb}\Sigma_{bb.a}A_{bb}^T = \Delta_{bb}(\Delta_{bb}^{-1} - \Theta_{a|b}^T\Delta_{aa-b}^{-1}\Theta_{a|b})\Delta_{bb}, \quad B_{bb}^T\Sigma_{bb.a}^{-1}B_{bb} = \Delta^{bb+a},$$

it follows in particular that both the conditional covariance matrix $\Sigma_{bb.a}$, as well as its inverse, can be expressed in terms of parameters (5) of system (4), with the inverse having a simpler representation, since Δ_{bb} is always a diagonal matrix, but Δ_{aa-b} is typically not of diagonal form.

Conversely, the marginal covariance matrix of Y_a has simpler representation than its inverse, which follows from the dual equalities

$$A_{aa}\Sigma_{aa}A_{aa}^T = \Delta_{aa-b}, \quad B_{aa}^T\Sigma_{aa}^{-1}B_{aa} = -\Delta_{aa}^{-1}(\Delta_{aa} - \Theta_{a|b}(\Delta^{bb+a})^{-1}\Theta_{a|b}^T)\Delta_{aa}^{-1}.$$

While proofs of the equalities are tedious by inverting special sums of matrices, they are direct by sweeping on the covariance and concentration matrix of $\varepsilon_{a-b} = A_{aa}Y_a$ and $\varepsilon_b = A_{bb}Y_b$, For system (4) these are given by

$$\mathrm{cov}(\varepsilon_{a-b}, \varepsilon_b) = \begin{pmatrix} \Delta_{aa-b} & -\Theta_{a|b}\Delta_{bb} \\ . & \Delta_{bb} \end{pmatrix}, \quad \mathrm{cov}(\varepsilon_{a-b}, \varepsilon_b)^{-1} = \begin{pmatrix} \Delta_{aa}^{-1} & \Delta_{aa}^{-1}\Theta_{a|b} \\ . & \Delta^{bb+a} \end{pmatrix},$$

respectively.

The linear equations corresponding to the conditional distribution of Y_b given Y_a derived from the recursive system of Figure 1 can then be written as

$$Y_b - \Pi_{b|a}y_a = B_{bb}(\varepsilon_b + \Pi_{\varepsilon_b|\varepsilon_{a-b}}\varepsilon_{a-b}), \quad Y_a = B_{aa}\varepsilon_{a-b} \qquad (11)$$

where $\Pi_{\varepsilon_b|\varepsilon_{a-b}} = (-\Delta_{bb}\Theta_{a|b}^T)\Delta_{aa-b}^{-1} = (\Delta^{bb+a})^{-1}(\Theta_{a|b}^T\Delta_{aa}^{-1})$ and $\varepsilon_{a-b} = \varepsilon_a - \Theta_{a|b}\varepsilon_b$, $\Pi_{b|a} = B_{bb}\Pi_{\varepsilon_b|\varepsilon_{a-b}}A_{aa}$, i.e. they are as defined above for (10).

Thus, when the order of dependencies is reversed as compared to the order in the generating process, it is possible, but not simple, to express the new mean parameters in terms of parameters of the generating process. This suggests that for general results on induced joint distributions it will be useful to preserve the original ordering as far as possible.

If the independence structure of Y_b given Y_a is expressed in terms of block-regression (Wermuth, 1992) and the independencies in the margin of Y_a as a concentration graph structure, then its representation is relatively simple despite the dependence reversal. *Block-regression* of Y_b on Y_a means that each single component of the response vector Y_b is regressed on all remaining components of this response vector and on all components of the explanatory vector variable Y_a so that an (i,j)-structural zero within b and between b and a means for joint Gaussian variables conditional independence given all remaining variables $(V \setminus \{i,j\})$.

Therefore, it follows with (5) that there is a structural zero in the block-regression equations of Y_b on Y_a if and only if there is a structural zero in $(A^TA)_{V,b}$, i.e. in the submatrix of all rows V and of columns b. The representation of Σ_{aa}^{-1} in (10) implies that there is an (i,j)-structural zero in the marginal concentration matrix of Y_a if and only if there is a (i,j)-structural zero in $A_{aa}^T\Delta_{aa-b}^{-1}A_{aa}$. These results are illustrated in Figure 3. We return to their interpretation in terms of graphs in Section 5.

Given just the independence structures of joint response models such as in Figures 2 and 3 it is in general not possible to recover the independence structure of the generating graph. Also, the independence structures derived from covariance graphs or concentrations graphs after marginalising or conditioning on some variables typically contain more edges than if they are derived directly from the generating graph. As we shall see this is different for the form of the independence structure derived in the next Section.

4. Linear recursive systems after marginalising and conditioning

Starting from the univariate generating process (4) we now divide the overall set of indices V into any three disjoint sets S, C, M and examine the distribution of Y_S given Y_C marginalising over Y_M.

Two main special cases were treated in Section 3 with Y_V partitioned into two ordered subvectors Y_a, Y_b, i.e. with $V = (a, b)$. If we condition Y_a only on potential ancestors Y_b as in (9) the distribution of Y_a given $Y_b = y_b$ is unchanged, it being determined by the subsystem $A_{aa}Y_{a|b} = \epsilon_a$, where $Y_{a|b}$ denotes the deviation of Y_a from its conditional mean given Y_b. If however we condition Y_b on common descendants within Y_a as in (11) the conditional distribution of Y_b given Y_a has correlated residuals. Correspondingly, if we marginalise only over descendants Y_a then the marginal distribution of Y_b is unchanged, whereas if we marginalise over common ancestors in Y_b the marginal distribution of Y_a has correlated residuals.

In general, however, there will be no simple ordering between the components forming S, C, M and a generalisation of the above arguments is required which we now set out.

4.1 Preliminaries, the two main results and the outline for deriving them

To keep the information on which components of S and which of $M = V \setminus (S \cup C)$ are ancestors of C in the generating directed acyclic graph we introduce first some further definitions. As mentioned before from an ancestor of C a direction-preserving path leads to a node in C. We refer to the nonancestors of a set C, i.e. to those nodes from which no direction-preserving path leads into C, as the *offspring* of C. For example, the offspring of node 3 in Figure 1 are the nodes in $\{1, 2, 4\}$; and, if we choose for Figure 1 the set of selected nodes as $S = \{1, 3, 4, 7, 10\}$ and the conditioning set as $C = \{6, 9\}$, then the remaining nodes are to be marginalised over, i.e. they are in $M = \{2, 5, 8\}$. For ancestors of C outside C we distinguish two possibilities. They are either in S and denoted by $S_{\text{anc}} = v$ or they are in M and denoted by $M_{\text{anc}} = q$. Correspondingly, there are offspring of C in S, denoted by $S_{\text{off}} = u$, and there are offpring of C in M, denoted by $M_{\text{off}} = p$. In this way

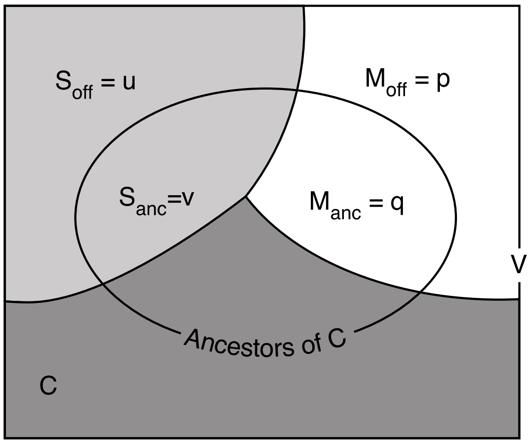


Figure 5: A partitioning of the node set to derive the structure in the distribution of Y_S given Y_C as induced by a generating directed acyclic graph in nodes V

S and M are further partitioned as $S = u \cup v$ and $M = p \cup q$. The definitions are illustrated with Figure 5. For the previous example to Figure 1 with $C = \{6, 9\}$, $S = \{1, 3, 4, 7, 10\}$ and $M = \{2, 5, 8\}$ we get $u = \{1, 3, 4\}$, $v = \{7, 10\}$ and $p = \{2, 5\}$, $q = \{8\}$.

As the first main result in this Section we derive the distribution of Y_S given Y_C as the conditional distribution of Y_u given both Y_C and Y_v, and the conditional distributon of Y_v given Y_C to be of the following form:

$$\Gamma_{uu} Y_{u|C} + \Gamma_{uv} y_{v|C} = \eta_u, \quad Y_{v|C} = \eta_v, \tag{12}$$

where $Y_{t|C}$ denotes a (vector) variable Y_t centred at its conditional mean given Y_C, Γ_{uu} and Γ_{vv} are upper-triangular matrices, the residuals between the two blocks, i.e. η_u and η_v, are uncorrelated, but residuals within each block may be correlated so that the system is still univariate recursive but possibly has equations in which one contains some information on the other.

As the second main result we prove that each system of connected univariate dependencies (12) has a covering model with the same structural zeros in the covariance matrix of Y_S given Y_C as implied by (12) but fewer structural zeros, i.e. fewer restrictions elsewhere. This covering model is simpler in structure than (12) since it is univariate recursive in the components of $Y_{V \setminus C}$ with exclusively independent residuals.

To derive the results we denote some further special subsets of all nodes $V = S \cup C \cup M$ as follows:

$$H = S_{\text{off}} \cup M_{\text{off}} = u \cup p, \quad D = S_{\text{anc}} \cup M_{\text{anc}} = v \cup q, \quad L = C \cup D, \quad K = C \cup v,$$

where offspring and ancestors are defined relative to the conditioning set C, as described for Figure 5. Offspring nodes H of C are also offspring of L and ancestors q of C outside C are also ancestors of K. The outline of the reasoning is then as follows. We get first the marginal distribution of Y_K (in 4.2), then the conditional distribution of Y_u given $Y_K = y_k$ (in 4.3). Finally, we condition both Y_u and Y_v on Y_C (in 4.4). As additional results we show (in 4.5) that marginalising first over Y_M, and conditioning next Y_S on Y_C leads to the same equations (12) and give (in 4.6) the concentration and the covariance matrix of Y_S given Y_C, both expressed in terms of parameters of (12) and of (4). Finally, the indicator matrix of structural zeros in the conditional covariance matrix of varibles outside C given C is derived via a covering model to (12) which is univariate recursive and has independent residuals.

Expressed more formally we shall derive the equations for the conditional distribution of Y_u given $Y_K = y_K$ and the marginal distribution of Y_K as

$$A_{uu.p}(Y_u - \Pi_{u|K} y_K) = \varepsilon_{u-M} - \Delta_{uK-M}\Delta^{-1}_{KK-q}\varepsilon_{K-q}, \quad A_{KK.q}Y_K = \varepsilon_{K-q}, \qquad (13)$$

where e.g. $\varepsilon_{K-q} = \varepsilon_K - A_{Kq}A^{-1}_{qq}\varepsilon_q$, we use the notation

$$\text{cov}(\varepsilon_{(V\setminus M)-M}) = \begin{pmatrix} \Delta_{uu-M} & \Delta_{uK-M} \\ . & \Delta_{KK-M} \end{pmatrix},$$

and note that the split of M into offspring p and ancestors q of C implies $\Delta_{KK-M} = \Delta_{KK-q}$ and $\Delta_{uu-M} = \Delta_{uu-p}$. Furthermore, $A_{aa.b}$ denotes the submatrix $(A^*)_{a,a}$ obtained after sweeping A on b and

$$A_{uu.p}\Pi_{u|K} = A_{uK.p} + A_{uq.p}\Pi_{q|K}, \quad A_{qq}\Pi_{q|K} = \Delta_{qq}B^T_{Kq}\Sigma^{-1}_{KK} - A_{qK}.$$

Thereby, we use for the matrix obtained after projecting Y_u linearly on Y_K, i.e. for $\Pi_{u|K}$, the matrix version of Cochran's (1938) recursion formula for linear regression coefficients: $\Pi_{u|K} = \Pi_{u|K.q} + \Pi_{u|q.K}\Pi_{q|K}$. It may be obtained with the help of the original sweep operator as shown here in Appendix 1 or by taking expectations of conditional expectations in equations regressing Y_u on Y_K. We partition the matrix $-\Pi_{u|L} = A^{-1}_{uu.p}A_{uL.p} = A^{-1}_{uu.p}(A_{uK.p} \;\; A_{uq.p})$ and note that the matrix of regression coefficients of Y_K has the special form $\Pi_{u|K.q} = -A^{-1}_{uu.p}A_{uK.p}$ and that, similarly, the one of Y_q is $\Pi_{u|q.K} = -A^{-1}_{uu.p}A_{uq.p}$.

To obtain the matrix $\Pi_{q|K}$ in terms of parameters of (4) the matrix products $A_{LL}\Sigma_{LL} = \Delta_{LL}B^T_{LL}$ are partitioned into K and q

$$\begin{pmatrix} A_{KK} & A_{Kq} \\ A_{qK} & A_{qq} \end{pmatrix} \begin{pmatrix} \Sigma_{KK} & \Sigma_{Kq} \\ \Sigma_{qK} & \Sigma_{qq} \end{pmatrix} = \begin{pmatrix} \Delta_{KK} & 0 \\ 0 & \Delta_{qq} \end{pmatrix} \begin{pmatrix} B^T_{KK} & B^T_{qK} \\ B^T_{Kq} & B^T_{qq} \end{pmatrix}.$$

The second row on the left multiplied by the first column with the matrix on the right gives $A_{qK}\Sigma_{KK} + A_{qq}\Sigma_{qK} = \Delta_{qq}B^T_{Kq}$. The matrix of regression coefficients $\Pi_{q|K} = \Sigma_{qK}\Sigma^{-1}_{KK}$ reduces to simpler forms, whenever nodes within q and K are adjacent. For instance if in that case descendants are conditioned on ancestors, so that $A_{Kq} = B_{Kq} = 0$, then $\Pi_{q|K} = -A^{-1}_{qq}A_{qK}$ as derived from (7) above.

4.2 The marginal distribution of Y_K

To illustrate first the result in (13) for the marginal distribution of an arbitrary subset K of L we partition L by preserving the order as $L = (a, b, c, d, e, f, g)$ and marginalise over $q = \{b, d, f\}$. Then the remaining nodes, still ordered as in the generating system, are $K = (a, c, e, g)$. We then get

$$A_{KK.q} = (A_{LL} \text{ swt } q)_{K,K} = \begin{pmatrix} A_{aa} & A_{ac.b} & A_{ae.bd} & A_{ag.bdf} \\ 0 & A_{cc} & A_{ce.d} & A_{cg.df} \\ 0 & 0 & A_{ee} & A_{eg.f} \\ 0 & 0 & 0 & A_{gg} \end{pmatrix}. \tag{14}$$

We note that

$$A^{-1}_{KK.q} = (A_{KK.q} \text{ swt } K) = (A_{LL} \text{ swt } q, K)_{K,K} = B_{KK},$$

and that as a consequence of $(A_{LL} \text{ swt } q) = (B_{LL} \text{ swt } K)$ the essential part of the residual $\varepsilon_{K-q} = \varepsilon_K - \Theta_{K|q}\varepsilon_q$ in the marginal distribution, i.e. $\Theta_{K|q}$, can be expressed in (at least) two different ways, as $\Theta_{K|q} = A_{Kq}A^{-1}_{qq} = -B^{-1}_{KK}B_{Kq}$ also when the indices within K and within q are not adjacent. For the above special choice of K and q this gives, written explicitly,

$$\Theta_{K|q} = (A_{LL} \text{ swt } q)_{K,q} = \begin{pmatrix} A_{ab} & A_{ad.b} & A_{af.bd} \\ 0 & A_{cd} & A_{cf.d} \\ 0 & 0 & A_{ef} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} A_{bb} & 0 & 0 \\ 0 & A_{dd} & 0 \\ 0 & 0 & A_{ff} \end{pmatrix}^{-1}, \tag{15}$$

and

$$\Theta_{K|q} = (B_{LL} \text{ swt } K)_{K,q} = \begin{pmatrix} B_{aa} & 0 & 0 & 0 \\ 0 & B_{cc} & 0 & 0 \\ 0 & 0 & B_{ee} & 0 \\ 0 & 0 & 0 & B_{gg} \end{pmatrix}^{-1} \begin{pmatrix} B_{ab} & B_{ad.c} & B_{af.ce} \\ 0 & B_{cd} & B_{cf.e} \\ 0 & 0 & B_{ef} \\ 0 & 0 & 0 \end{pmatrix}. \tag{16}$$

Several useful further matrix equalities result from $(A_{LL} \text{ swt } q)^{-1} = (A_{LL} \text{ swt } K) = (B_{LL} \text{ swt } q)$ written for a partitioning of L into K and q as

$$\begin{pmatrix} A_{KK.q} & A_{Kq}A_{qq}^{-1} \\ -A_{qq}^{-1}A_{qK} & A_{qq}^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} A_{KK}^{-1} & -A_{KK}^{-1}A_{Kq} \\ A_{qK}A_{KK}^{-1} & A_{qq.K} \end{pmatrix} = \begin{pmatrix} B_{KK.q} & B_{Kq}B_{qq}^{-1} \\ -B_{qq}^{-1}B_{qK} & B_{qq}^{-1} \end{pmatrix}.$$

For instance, from the product of the first row on the left with the first column of the matrix in the middle it follows that $A_{KK.q} = A_{KK} - A_{Kq}A_{qq}^{-1}A_{qK}$, i.e. it has the same form as in the case when the indices are adjacent, and from the product of the first row on the left with the second column of the matrix on the right it follows that $B_{Kq} = -A_{KK.q}^{-1}A_{Kq}A_{qq}^{-1}$.

For completeness we note the form of the remaining parts of A_{LL} after sweeping it on q

$$(A_{LL} \text{ swt } q)_{q,K} = \begin{pmatrix} A_{bb} & 0 & 0 \\ 0 & A_{dd} & 0 \\ 0 & 0 & A_{ff} \end{pmatrix}^{-1} \begin{pmatrix} 0 & A_{bc} & A_{be.d} & A_{bg.df} \\ 0 & 0 & A_{de} & A_{dg.f} \\ 0 & 0 & 0 & A_{fg} \end{pmatrix}, \tag{17}$$

and

$$A_{qq}^{-1} = \begin{pmatrix} A_{bb} & 0 & 0 \\ 0 & A_{dd} & 0 \\ 0 & 0 & A_{ff} \end{pmatrix}^{-1} \begin{pmatrix} A_{bb} & -A_{bd} & -A_{bf.d} \\ 0 & A_{dd} & -A_{df} \\ 0 & 0 & A_{ff} \end{pmatrix} \begin{pmatrix} A_{bb} & 0 & 0 \\ 0 & A_{dd} & 0 \\ 0 & 0 & A_{ff} \end{pmatrix}^{-1}. \tag{18}$$

In spite of the apparent similarities in form of the entries in the 3×4 block matrix obtained from the right-hand matrix product in (17) are in general not matrices of regression coefficients.

The marginal covariance and concentration matrix of Y_K may now be written as

$$\mathrm{cov}(Y_K) = \Sigma_{KK} = B_{KK}\Delta_{KK-q}B_{KK}^T, \quad \mathrm{con}_K(Y_K) = \Sigma_{KK}^{-1} = A_{KK.q}^T\Delta_{KK-q}^{-1}A_{KK.q}. \tag{19}$$

There is a structural zero in Δ_{KK-q} if and only if there is a zero element in the symbolic outer matrix product $\Theta_{K|q}\Theta_{K|q}^T$. For structural zeros in Δ_{KK-q}^{-1} the explanation is different. It depends crucially on q containing exclusively ancestors of C outside the set C as is explained in the following.

For general covariance matrices some nonzero entries may vanish after conditioning. For instance, for Figure 1, the marginal covariance corresponding to nodes (6,9) is not a structural zero, but after conditioning on 8 or 10 it becomes a structural zero. However, for the residual covariance matrix Δ_{KK-q} every entry corresponding to a node pair (i, j) within K is a free parameter if and only if either Y_j is a directly explanatory variable for Y_i in the generating system (4), or the pair has a common explanatory variable within q such that all intermediate variables are also in q. Such a free parameter cannot turn into a structural zero by conditioning on nodes of $K = L \setminus q$.

This implies in particular that Δ_{KK-q}^{-1} has a free (i, j)-parameter if and only if there is a sequence of nodes $i = i_0, i_1, \ldots, i_t, i_{t+1} = j$ such that each adjacent pair corresponds to a free parameter in Δ_{KK-q}, otherwise it has a structural (i, j)-zero.

Alternatively, structural zeros in the concentration matrix of Y_K may be expressed in terms of parameters of the generating system (4) as follows. In general concentration matrices some nonzero entries can vanish after marginalising. For instance in Figure 2, the concentration corresponding to nodes (7,8) is not a structural zero, but after marginalising over node 6 it becomes a structural zero. However, this cannot happen if marginalising is exclusively over ancestor nodes of a given conditioning set C not in C. This is the case when $K = C \cup v$ is obtained from L by marginalising over q, since by definition every node in q is an ancestor of C outside C.

With $L = V \setminus H = K \cup q$ denoting adjacent indices, it follows from (5) that there is a structural zero in $\mathrm{con}_L(Y_K)$ if and only if there is one in the matrix product

$A_{LL}^T A_{LL}$. By marginalising L over ancestors q of C outside C some structural zeros present in $\text{con}_L(Y_K)$ can get removed but none can get added in $\Sigma_{KK}^{-1} = \text{con}_K(Y_K)$. Inverting a concentration matrix corresponds to marginalising. If in the conditional covariance matrix of Y_q given Y_C there is a sequence of free parameters an independence statement for the endpoints is no longer preserved after marginalising over nodes q.

From the matrix version of Dempster's (1969) recursion formula for concentrations, $\Sigma^{aa.bc} = \Sigma^{aa.c} - \Sigma^{ab.c}(\Sigma^{bb.c})^{-1}\Sigma^{ba.c}$, it then follows in the special case of interest here, i.e. for $a = K,\ b = q,\ c = H$, that there is a (i,j)-structural zero in $\text{con}_K(Y_K)$ if and only if there is such a structural zero in $\text{con}_L(Y_K)$ and there is no sequence $i = i_0, i_1, \ldots, i_t, i_{t+1} = j$ with indices $i_l,\ l = 1, \ldots, t$ within q such that each adjacent index pair in the whole sequence corresponds to a free parameter in $\text{con}_L(Y_K)$.

In preparation for the next sections we take as one example to Figure 1 $v = \{2, 9\}$ and $C = \{1, 4, 6\}$ so that the set H of offspring of C is empty and $q = \{3, 5, 7, 8, 10\}$. Figure 6 shows the overall concentration graph with nodes of $K = C \cup v$ being darkened. It has an edge present if and only if there is a free parameter in $\text{con}_V(Y) = A^T\Delta^{-1}A$. No structural zero remains after marginalising over q in Figure 6, that is in $\text{con}_K(Y_K)$, since for every unconnected pair (i,j) within K there is in this example a sequence of free parameters in q connecting i and j.

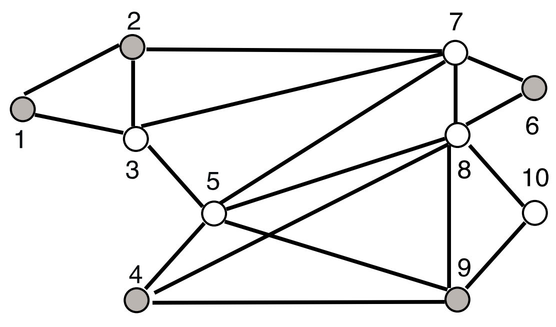


Figure 6: The overall concentration graph induced by the generating graph in Figure 1, nodes for $K = \{1, 2, 4, 6, 9\}$ are darkened; marginalising over the remaining nodes which in Figure 1 are all ancestors of K removes each structural zero within K

As another example to Figure 1 let $v = \{7, 8, 9\}$, $u = \emptyset$, $C = \{1, 2, 3, 5\}$, and $q = 10$, so that $H = p = \{4, 6\}$ and $L = V \setminus H$ consists of the remaining eight nodes.

Then the submatrix of v in $\mathrm{con}_L(Y_L)$ is diagonal, i.e. has three structural zeros, while the structural zero for (8,9) is removed after marginalising over 10. Thus, a structural (i,j)-zero present in A for ancestors v of a conditioning set C can be removed in $\mathrm{con}_K(Y_K)$ by conditioning on K, by marginalising over q or only by a combination of the two.

4.3 The conditional distribution of Y_u given Y_K

To obtain the special form in equations (13) for the conditional distribution of Y_u given Y_K we note first that by sweeping A on p at most components corresponding to the offspring H of C are modified but not those in $V \setminus H = L$. In particular, after writing $(A \text{ swt } p)_{u,V\setminus p} = (A_{uu.p} \;\; A_{uL.p})$ we get the corresponding linear equations for the joint distribution of Y_u, Y_L as

$$A_{uu.p}Y_u + A_{uL.p}y_L = \varepsilon_{u-p}, \quad A_{LL}Y_L = \varepsilon_L.$$

In these equations an essential part of the order of the variables in the generating system is preserved, since u consists of offspring of C and hence of $L = C\cup$ ancestors of C. Indices within L are adjacent, but within u they need not be. The residuals between the two blocks are uncorrelated, since they have no components in common. The residuals within u may be correlated and the matrices $A_{uu.p}$ and A_{LL} are both of upper triangular form. Therefore, the conditional distribution of Y_u given Y_L is described by equations which represent univariate recursive regressions conditionally given Y_L and have possibly connected dependencies within block u. The matrix of regression coefficients in the linear regression of Y_u and Y_L is $\Pi_{u|L} = -A_{uu.p}^{-1}A_{uL.p}$ since ε_{u-p} is uncorrelated with ε_L and hence with Y_L.

After partitioning L into q and K, moving $A_{uq.p}Y_q$ to the right-hand side and adding $A_{uq.p}\Pi_{q|K}Y_K$ on both sides, the equations for Y_u given Y_L are modified into equations of Y_u given Y_K alone as

$$A_{uu.p}(Y_u - \Pi_{u|K}y_K) = \varepsilon_{u-p} - A_{uq.p}(Y_q - \Pi_{q|K}y_K). \tag{20}$$

Since nodes q and K need not be adjacent we have $Y_q = A_{qq}^{-1}(\varepsilon_q - A_{qK}Y_K)$ and since, as explained above, $A_{qq}\Pi_{q|K} = \Delta_{qq}B_{Kq}^T\Sigma_{KK}^{-1} - A_{qK}$, it follows that

$$Y_q - \Pi_{q|K}y_K = A_{qq}^{-1}(\varepsilon_q - \Delta_{qq}B_{Kq}^T\Sigma_{KK}^{-1}y_K).$$

For the equivalent form of (20) to the first equations in (13) we need further that $\varepsilon_{u-M} = \varepsilon_{u-p} - A_{uq.p}A_{qq}^{-1}\varepsilon_q$, $\Delta_{uK-M} = A_{uq.p}A_{qq}^{-1}\Delta_{qq}A_{qq}^{-T}A_{Kq}^T$, $B_{Kq}^T = -A_{qq}^{-T}A_{Kq}^{-T}A_{KK.q}^{-T}$, $\Sigma_{KK}^{-1} = A_{KK.q}^T\Delta_{KK-q}^{-1}A_{KK.q}$ and $Y_K = A_{KK.q}^{-1}\varepsilon_{K-q}$.

From the explicit form for the residuals in (20) it can be derived directly that they have zero covariance with the residuals ε_{KK-q}. Therefore, the equations in u of (20) for selected offspring of the conditioning set C remain unchanged when the joint marginal distribution of C and selected ancestors v is represented in different ways, for instance by the covariance matrix, by the concentration matrix or by the univariate recursive system of connected dependencies $A_{KK.q}Y_K = \varepsilon_{K-q}$ of Section 4.2.

4.4 Deriving equations and structural zeros for the distribution of $Y_{S|C}$

Equations (13) specify a linear system for the joint distribution of Y_u and Y_K in two blocks which have uncorrelated residuals between blocks. Equations for the joint distribution of Y_S given Y_C are to be derived from them, where $S = u \cup v$ and $K = C \cup v$. This is achieved by rewriting equations (13) explicitly in terms of v and C as

$$\Gamma_{uu}Y_u + \Gamma_{uv}y_v + \Gamma_{uC}y_C = \eta_u, \quad \Gamma_{KK}Y_K = \eta_K.$$

The concentration matrix of the residuals after regressing Y_v on Y_C, i.e. after taking $Y_v = \Pi_{v|C}y_C + \varepsilon_{v|C}$, is the (v, v)-submatrix of $\Sigma_{KK}^{-1} = \mathrm{con}_K(Y_K)$. The matrix of regression coefficients obtained by projecting Y_v on Y_C, i.e. $\Pi_{v|C}$ can for example be obtained by sweeping Σ_{KK}^{-1} on v, but explicit expressions are complex.

After writing $Y_{v|C} = Y_v - \Pi_{v|C}y_C$ and inserting y_v in the first equation we get

$$\Gamma_{uu}Y_u + \Gamma_{uv}(\Pi_{v|C}y_C + y_{v|C}) + \Gamma_{uC}y_C = \eta_u,$$

and observe that η_u is uncorrelated with $y_{v|C}$ since it is uncorrelated with y_K. Finally, we obtain from the explicit form of the Γ-matrices in (13) that

$$Y_u + (\Gamma_{uu})^{-1}(\Gamma_{uC} + \Gamma_{uv}\Pi_{v|C})y_C = Y_u - (\Pi_{u|C.v} + \Pi_{u|v.C}\Pi_{v|C})y_C = Y_u - \Pi_{u|C}y_C.$$

so that the joint distribution has the claimed form (12)

$$A_{uu.p}Y_{u|C} + A_{uu.p}\Pi_{u|v.C}y_{v|C} = \eta_u, \quad Y_{v|C} = \eta_v.$$

With $\Pi_{u|v.C}\Sigma_{vv.C} = \Sigma_{uv.C}$ the covariance matrix of the residuals results as

$$\text{cov}(\eta) = \begin{pmatrix} \Psi_{uu} & 0 \\ 0 & \Psi_{vv} \end{pmatrix} = \begin{pmatrix} A_{uu.p}\Sigma_{uu.vC}A_{uu.p}^T & 0 \\ 0 & \Sigma_{vv.C} \end{pmatrix},$$

and, as mentioned above, $\Sigma_{vv.C}^{-1} = (\Sigma_{KK}^{-1})_{v,v}$. This completes the proof of (12).

4.5 Orthogonalising correlated residuals

To understand how marginalising first over $M = p \cup q$ and then conditioning Y_u on Y_K leads to the same form of equations as given in (13) and (12) we note that

$$(A \text{ swt } M)_{V\setminus M, V\setminus M} = \begin{pmatrix} A_{uu.p} & A_{uK.pq} \\ 0 & A_{KK.q} \end{pmatrix}, \quad \varepsilon_{u-M} = \varepsilon_{u-p} - A_{uq.p}A_{qq}^{-1}\varepsilon_q.$$

Since ε_{u-M} and ε_{K-q} are in general correlated, because both contain ε_q, an orthogonalisation step is needed to get from

$$A_{uu.p}Y_u + A_{uK.pq}y_K = \varepsilon_{u-M}, \quad A_{KK.q}Y_K = \varepsilon_{K-q}$$

to the equations in (13). This is achieved by subtracting $\Pi_{\varepsilon_{u-M}|\varepsilon_{K-q}}A_{KK.q}Y_K$ from both sides of the equations, observing that

$$\Pi_{\varepsilon_{u-M}|\varepsilon_{K-q}} = \Delta_{uK-M}\Delta_{KK-q}^{-1}, \quad A_{uK.pq} = A_{uK.p} - A_{uq.p}A_{qq}^{-1}A_{qK}$$

and making again use of some of the above matrix equalities.

4.6 The covariance and concentration matrices of Y_S given Y_C

With the explicit results for the equations of Y_u given Y_C, Y_v and of Y_v given Y_C, at the end of Section 4.4, the conditional covariance and concentration matrix of Y_S given Y_C are directly expressible in terms of parameters of the system of univariate connected dependencies (12).

The covariance matrix of Y_S given Y_C is

$$\Sigma_{SS.C} = \begin{pmatrix} \Gamma_{uu}^{-1} & -\Gamma_{uu}^{-1}\Gamma_{uv} \\ 0 & \mathrm{I}_{vv} \end{pmatrix} \begin{pmatrix} \Psi_{uu} & 0 \\ 0 & \Psi_{vv} \end{pmatrix} \begin{pmatrix} \Gamma_{uu}^{-T} & 0 \\ -\Gamma_{uv}^{T}\Gamma_{uu}^{-T} & \mathrm{I}_{vv} \end{pmatrix}, \qquad (21)$$

where I_{vv} denotes the identity matrix of a size corresponding to Y_v, $\Gamma_{uu} = A_{uu.p}$ and $-\Gamma_{uu}^{-1}\Gamma_{uv} = \Pi_{u|v.C}$. There can be no additional structural zero in $\Psi_{vv} = \Sigma_{vv.C}$ which is not present in its inverse Ψ_{vv}^{-1}, since an additional structural (i,j)-zero in v could only be generated by marginalising over an index h within v which has both i and j as parents in the generating process but no descendant in C. Since v contains only ancestors of C this would contradict the definition. Hence, structural zeros can only get removed by inverting Ψ_{vv}^{-1}, i.e. by marginalising over nodes of v. More precisely, there is a (i,j)-structural zero in Ψ_{vv} if and only if there is such a structural zero in Ψ_{vv}^{-1} and there is no sequence $i = i_0, i_1, \ldots, i_t, i_{t+1} = j$ with indices $i_l,\ l = 1, \ldots, t$ within v such that each adjacent index pair in the whole sequence corresponds to a free parameter in Ψ_{vv}^{-1}.

This implies in particular that Ψ_{vv} has a *complete block-diagonal form*, i.e. it consists exclusively of complete nonoverlapping blocks. Blocks indicate (vector) components of v which are mutually independent and remain independent after conditioning on C.

The concentration matrix of Y_S given Y_C is then

$$\Sigma_{SS.C}^{-1} = \begin{pmatrix} \Gamma_{uu}^{T}\Psi_{uu}^{-1}\Gamma_{uu} & \Gamma_{uu}^{T}\Psi_{uu}^{-1}\Gamma_{uv} \\ . & \Psi_{vv}^{-1} \end{pmatrix} = \begin{pmatrix} \Sigma^{uu.M} & \Sigma^{uv.M} \\ . & \Sigma^{vv.M} \end{pmatrix},$$

where the right-hand side shows the notation after sweeping the overall concentration matrix on M, i.e. for marginalising in Σ^{-1} over M. There is for instance a structural (j,k)-zero in $\Gamma_{uu}^{T}\Psi_{uu}^{-1}\Gamma_{uv}$ if and only if each individual term in the matrix product is zero, i.e. if and only if $\gamma_{hj}(\Psi_{uu}^{-1})_{hi}\gamma_{ik} = 0$ for indices $h, i, j \in u$ and $k \in v$. By conditioning on u, i.e. by inverting Ψ_{uu}, no additional structural zeros can be induced since an edge in it is present in the generating graph or it is generated by marginalising over M or conditioning on K, hence it cannot be removed by conditioning on indices outside M and K.

For the representation of $\mathrm{con}_{V\setminus M}(Y_S, Y_C)$ in terms of the parameters of the generating system (4), we partition M into r and w, where w are ancestors of $S \cup C$ outside $S \cup C$ and r are offspring of $S \cup C$ and denote the union of S, C, w by $N = V \setminus r$. Marginalising over offspring leaves the marginal distibution of the remaining variables unchanged. Therefore the concentration marix of Y_N is $\Sigma^{NN.r} = A^T_{NN}(\Delta_{NN})^{-1}A_{NN}$. There are typically more structural zeros in $\Sigma^{NN.r}$ than in the submatrix of N in the overall concentration graph, i.e. in $(\Sigma^{-1})_{N,N}$. By marginalising over the ancestors w of $S \cup C$ no additional structural zero can get induced but only structural zeros present in $\Sigma^{NN.r}$ can get removed.

These properties assure that in

$$\mathrm{con}_{V\setminus M}(Y_S, Y_C) = \begin{pmatrix} \Sigma^{SS.r} & \Sigma^{SC.r} \\ . & \Sigma^{CC.r} \end{pmatrix} - \begin{pmatrix} \Sigma^{Sw.r} & \Sigma^{Cw.r} \end{pmatrix} (\Sigma^{ww.r})^{-1} \begin{pmatrix} \Sigma^{wS.r} \\ \Sigma^{wC.r} \end{pmatrix}$$

there is a structural (i, j)-zero if and only if there is one in $\Sigma^{NN.r}$ and there is no sequence of indices $i = i_0, i_1, \ldots, i_t, i_{t+1} = j$ with indices i_l, $l = 1, \ldots, t$ within w such that each adjacent index pair in the whole sequence corresponds to a free parameter in $\Sigma^{NN.r}$. This applies in particular for the submatrix of interest, i.e. for

$$\Sigma^{-1}_{SS.C} = (\mathrm{con}_{V\setminus M}\{Y_S, Y_C\})_{S,S}.$$

To express the covariance matrix of Y_S given Y_C in terms of the parameters of the generating system (4), i.e. for $AY = \varepsilon$, it is useful to partition the set of all indices V into the three components H, C, D, with C the conditioning set, D the ancestors of C outside C, and H offspring of C, and to use

$$\Sigma_{SS.C} = (\Sigma \text{ swp } C)_{S,S} = (-\Sigma^{-1} \text{ rswp } H, D)_{S,S}.$$

Now, the conditional concentration matrix of D given C is obtained after marginalising the overall concentration matrix over H and it is also the submatrix of the concentration matrix of $L = C \cup D$, i.e.

$$\Sigma^{-1}_{DD.C} = \Sigma^{DD.H} = (A^T_{LL}\Delta^{-1}_{LL}A_{LL})_{D,D}.$$

Since $\Sigma^{-1}_{DD.C}$ contains only ancestors of C the type of reasoning given before applies here as well, i.e. there is an (i, j)-edge in its inverse $\Sigma_{DD.C}$ if and only if there is a sequence of free parameters between i and j in $\Sigma^{-1}_{DD.C}$.

After resweeping the overall concentration matrix on H, the submatrix for $V \backslash C$ is

$$\begin{pmatrix} \Sigma_{HH.CD} & \Pi_{H|D.C} \\ . & -\Sigma^{DD.H} \end{pmatrix} = \begin{pmatrix} (A_{HH}^T \Delta_{HH}^{-1} A_{HH})^{-1} & -A_{HH}^{-1} A_{HD} \\ . & -\Sigma_{DD.C}^{-1} \end{pmatrix},$$

and resweeping further on D leads to

$$\Sigma_{HD.C} = \Pi_{H|D.C} \Sigma_{DD.C}, \quad \Sigma_{HH.C} = B_{HH} \Delta_{HH} B_{HH}^T + \Pi_{H|D.C} \Sigma_{DD.C} \Pi_{H|D.C}^T.$$

In the special case when indices in C and D are adjacent, we have in the notation of Section 3 $\Sigma_{DD.C}^{-1} = A_{DD}^T \Delta^{DD+C} A_{DD}$ and $\Sigma_{HD.C} = B_{HD}(\Delta^{DD+C})^{-1} B_{DD}^T$.

The linear equations obtained after only conditioning in system (4) on an arbitrary subset C can be therefore be written in the form of (12) as

$$A_{HH} Y_{H|(C,D)} = A_{HH} Y_{H|C} + A_{HD} y_{D|C} = \varepsilon_H, \quad Y_{D|C} = \eta_D, \tag{22}$$

where the residuals ε_H have a diagonal covariance matrix and are uncorrelated with η_D. With these equations the conditional covariance matrix of nodes outside C given C, $\Sigma_{(V \backslash C, V \backslash C).C}$, is expressible as a special case of (21). It contains $\Sigma_{SS.C}$ as a submatrix and it can be rewritten as

$$\Sigma_{(V \backslash C, V \backslash C).C} = \begin{pmatrix} A_{HH}^{-1} & \Pi_{H|D.C} F_{DD} \\ . & F_{DD} \end{pmatrix} \begin{pmatrix} \Delta_{HH} & 0 \\ . & \Delta_{DD}^{+} \end{pmatrix} \begin{pmatrix} A_{HH}^{-T} & 0 \\ F_{DD}^T \Pi_{H|D.C}^T & F_{DD}^T \end{pmatrix},$$

where Δ_{DD}^{+} is a triangular matrix, F_{DD} is upper triangular and $(F_{DD}, \Delta_{DD}^{+})$ is a triangular decomposition of $\Sigma_{DD.C}$. We denote by $\mathcal{F}$ the indicator matrix of structural zeros in this triangular decomposition (F, Δ^{+}) of $\Sigma_{(V \backslash C, V \backslash C).C}$ and by $\mathcal{F}^{+}$ the indicator matrix of structural zeros in $\Sigma_{(V \backslash C, V \backslash C).C}$ itself, i.e. in FF^T.

Since the covariance matrix of the ancestors of C outside C is of complete block-diagonal form, the structural zeros in F_{DD} coincide with those in the outer product $F_{DD} F_{DD}^T$, i.e. $\mathcal{F}_{DD} = \mathcal{F}_{DD}^{+}$, and they remain unchanged for any order in which the triangular decompostion is carried out. Thus, the conditional covariance matrix of ancestors of a conditioning set as implied by system (4) has always a triangular decomposition of complete block-diagonal form which reflects all its structural zeros. But, the inverse of F_{DD} is in general not complete block-diagonal.

Also, the analogous statement is not true in general for the triangular decomposition of the concentration matrix $\Sigma_{DD.C}^{-1}$ as implied by system (4), i.e. the structural zeros in F_{DD}^{-1} may differ from those in the inner product $F_{DD}^{-T}F_{DD}^{-1}$. The reason is as follows. By conditioning on common responses in a univariate recursive system undirected chordless n-cycles can be generated in the corresponding overall concentration graph. In that case there exists no triangular decomposition of $\Sigma_{DD.C}^{-1}$ which has the same zero pattern as the upper triangular part of this concentration matrix, no matter which ordering of the variables is chosen (Wermuth, 1980; Cox & Wermuth, 1999).

We use these properties to define a special system of univariate recursive regressions in uncorrelated residuals for every conditional distribution of $Y_{V\setminus C}$ given Y_C. We introduce a matrix E such that

$$EY_{(V\setminus C)|C} = \begin{pmatrix} A_{HH} & A_{HD} \\ 0 & E_{DD} \end{pmatrix} \begin{pmatrix} Y_{H|C} \\ Y_{D|C} \end{pmatrix} = \begin{pmatrix} \varepsilon_H \\ \varepsilon_D^* \end{pmatrix} = \varepsilon^*,$$

where $\mathcal{E}_{DD} = \mathcal{F}_{DD} = \mathcal{F}_{DD}^{+}$, i.e. the subsystem for the ancestors D of C is block-diagonal and it has a structural zero whenever $\Sigma_{DD.C}$, the conditional covariance matrix implied by (4) for the ancestors of C has a structural zero. The equations for the offspring H of C are those of (4), just rewritten for Y_H in deviation from its conditional mean given C.

For any recursive system of complete block-diagonal form, such as $E_{DD}Y_{D|C} = \varepsilon_D^*$, the indicator matrices of structural zeros coincide for the defining upper triangular matrix (E_{DD}), its inverse, and their inner and outer products. This means that we have also chosen E_{DD} so that $\Sigma_{DD.C}^{*} = E_{DD}^{-1}\mathrm{cov}(\varepsilon_{DD}^{*})E_{DD}^{-1}$ has the same structural zeros as its inverse. Furthermore, since

$$E^{-1} = \begin{pmatrix} A_{HH}^{-1} & -A_{HH}^{-1}A_{HD}E_{DD}^{-1} \\ 0 & E_{DD}^{-1} \end{pmatrix}$$

and the structural zeros in E_{DD} coincide with those of E_{DD}^{-1} it follows that the indicator matrix of structural zeros in E^{-1} is identical to $\mathcal{F}$ and that therefore the structural zeros in $\Sigma_{(V\setminus C,V\setminus C).C} = F\Delta^{+}F^{T}$ agree with those in $\Sigma_{(V\setminus C,V\setminus C).C}^{*} = E^{-1}\mathrm{cov}(\varepsilon^*)E^{-T}$.

This is a further example of a covering model which is simpler to analyze than the reduced model embedded in it (Cox & Wermuth, 1990). Here the covering model is univariate recursive with independent residuals, the reduced model (22) need not have such a representation which reflects its independence structure fully. The additional restrictions in the reduced model are independencies for components of $Y_{D|C}$ with a conditioning set larger than C. They correspond to structural zeros in the inverse of $\Sigma_{DD.C}^{-1}$ not present in $\Sigma_{DD.C}$ as well. We have shown these independencies are not needed to decide on the missingness of an edge in the conditional covariance matrix, i.e. on whether the conditional independence $Y_i \perp\!\!\!\perp Y_j \mid Y_C$ is implied by the generating system (4).

Generalizations to other than linear systems are studied in detail next. In particular we prove in Section 5.9 the equivalence of matrix conditions for constructing the conditional covariance graph of Y_S given Y_C to a simple path criterion and show their validity for general distributions factorizing as in (1), i.e. as determined by the parent graph. Finally in Section 5.10 a programmable matrix algorithm is provided to obtain the edge matrix of a conditional covariance graph directly from the edge matrix of the generating parent graph.

5. Generalizations to arbitrary distributions generated over graphs

We have seen in Section 4 that a linear recursive system (4) with uncorrelated residuals can be turned after marginalising and conditioning into univariate recursive regressions having typically some correlated residuals for offspring of the conditioning set and, independently, into a covariance selection model for ancestors of the conditioning set. That is, if we start with S, C, M as any disjoint subsets of the set of variables V in the generating system and $AY = \varepsilon$, where A is constrained only by having zeros whenever an edge is missing in the generating graph and ε is any zero mean vector with diagonal covariance matrix, then every conditional distribution of Y_S given Y_C marginalising over Y_M is of the form (12) for which we define associated graphs below.

These results can be generalized to systems in which responses, intermediate and explanatory variables may be discrete or continuous and the form of the joint distribution is arbitrary except that it is generated as described in Section 1 over a directed acyclic graph with given edge matrix $\mathcal{A}$ so that the joint density factorizes as in (1). Effects of marginalizing and conditioning have been studied from different perspectives by Koster (1999b), Richardson & Spirtes (2000), Wermuth et al. (2000).

One key used here is to establish the relation of forming inner and outer products of matrices and of sweeping triangular matrices to completing certain 3-node-configurations in graphs. Therefore we give in this Section first definitions of special type of V-configurations, paths and graphs. Next we restate the conditions for structural zeros in matrices of linear systems obtained from (4) in Section 4 together with equivalent conditions for missing edges in graphs and derive graphs with identical edge matrices by completing V-configurations, that is without reference to linear systems.

The conditions on structural zeros in matrices translate into factorizations of densities present in (1) being preserved after marginalising and conditioning. As the main result we shall see that if in a linear system generated over a parent graph of A a structural zero in matrices resulting from A after marginalising and conditioning implies for all Gaussian distributions a conditional independence statement, then that same conditional independence holds for arbitrary distributions generated over the same parent graph.

5.1 Types of V-configuration, path and graph

Subgraphs induced by three nodes in a given graph in which two edges are present and one is absent are called V-*configurations*. The types of V- configurations in a directed acyclic graph differ in the configurations of the arrows at the common neighbour node t. A V-configuration is called *collision-oriented, transition-oriented* or *source-oriented*, respectively, depending on whether the common neighbour t is

$$i \longrightarrow t \longleftarrow j \qquad\qquad i \longleftarrow t \longleftarrow j \qquad\qquad i \longleftarrow t \longrightarrow j$$

a sink or collision node (left), a transition node (middle), or a source node (right).

With several paths passing through a node, this node may take on different roles along the different paths.

It is useful to characterize some further special type of paths in graphs of joint response models which we define below and study in the following Sections. As mentioned before, a path is said to be direction-preserving whenever it consists exclusively of arrows pointing in the same direction. A path is said to be a *pure collision path* if for every node along it has one of the following three configurations

$$\boldsymbol{i} \longrightarrow \boldsymbol{t} \longleftarrow \boldsymbol{j} \qquad \boldsymbol{i} - - - \boldsymbol{t} \longleftarrow \boldsymbol{j} \qquad \boldsymbol{i} - - - \boldsymbol{t} - - - \boldsymbol{j}$$

where the roles of i and j may also be reversed in the middle one and dashed lines are edges in covariance graphs (Cox & Wermuth, 1993, Wermuth & Cox, 1998). Pure collision paths of more than three nodes may differ in the type of the first and the last edge. Otherwise, they all have dashed lines along the path. If they did not, a noncollision node would occur along the path, contradicting the definition.

A path is *collisionless* if it does not contain any one of the above three configurations. In directed acyclic paths a collisionless path is either direction-preserving or it is a *common-source path*, i.e it consists of two direction-preserving paths, where the direction of the arrows changes at the common source node. In the other graphs to be derived here three further collisionless V-configurations are possible:

$$\boldsymbol{i} \longleftarrow \boldsymbol{t} \text{ ------ } \boldsymbol{j} \qquad \boldsymbol{i} \longleftarrow \boldsymbol{t} - - - \boldsymbol{j} \qquad \boldsymbol{i} \text{ ------ } \boldsymbol{t} \text{ ------ } \boldsymbol{j}$$

where full lines are edges in concentration graphs (Cox & Wermuth, 1993, Wermuth & Cox, 1998).

The different types of graph that we study here are all induced by a directed acyclic graph, G^V_{dag} with a given ordering of all nodes. We say they are induced by a given parent graph, G^V_{par} of A, where node 1 corresponds to row one and node d_V to the last row of A. If each nonzero (i, j)-entry of A implies a nonvanishing dependence of Y_i on Y_j given $Y_{\text{par}(i)\setminus j}$ then the joint distribution generated must have the global factorization property (1), as defining independence structure (2), and no additional independencies hold except those implied by (1) and (2) unless there are *parametric cancellations*, i.e. unless there are special parametric constellations. Conditions for the absence of such cancellations have been studied (Wermuth and Cox, 1998) but are of no relevance for the presence of structural zeros.

The specific types of graph to be derived from G^V_{par} of A are its ancestor graph, covariance and concentration graphs, graphs of connected univariate dependencies, and graphs of projecting one vector variable on another. The following list of definitions of graphs are related in the next Sections to structural zeros in transformations of the matrix A of a corresponding linear system. In graphs of connected dependencies an edge may have two components, in all other graphs each edge is simple, i.e. has just one component. In all graphs introduced here there are no directed cycles, i.e. it is impossible to follow the arrows along a direction-preserving path and return to the node from which one had started. As before we take the distinct nodes (h, i, j, k) to be in increasing order.

A generating process with independence structure (2) implies that the (i, j)-arrow is missing in G^V_{par} of A if and only if for $i, j \in V$

$$i \perp\!\!\!\perp j \mid \text{ parents of } i.$$

The *q-line ancestor graph*, $G^V_{\text{anc}(q)}$ of A, is a fully directed graph. It has an (i, j)-arrow present if and only if in G^V_{par} of A node j is a parent of i or an ancestor of i with all nodes along the path in q. In general a missing edge in this ancestor graph need not correspond to an independence statement. But in the overall ancestor graph an (i, j)-arrow is missing if and only if for $i, j \in V$

$$i \perp\!\!\!\perp j \mid \text{ potential ancestors of } j.$$

The *covariance graph of* Y_S given Y_C, $G^{S.C}_{\text{cov}}$ of A, is an undirected graph of dashed lines. The (i, j)-dashed-line is missing in it if and only if the generating process implies for $i, j \in S$

$$i \perp\!\!\!\perp j \mid C.$$

The *concentration graph of* Y_S given Y_C, $G^{S.C}_{\text{con}}$ of A, is an undirected graph of full lines. The (i, j)-full-line is missing in it if and only if the generating process implies for $i, j \in S$

$$i \perp\!\!\!\perp j \mid C \cup S \setminus \{i, j\}.$$

In $G^{S.C}_{\text{cud}}$ of A, the graph of *connected univariate dependencies* of Y_S given Y_C, there can be three types of edge, full lines, dashed lines or arrows. The selected nodes S are partitioned into u containing offspring of C and into v containing ancestors of C outside C. The subgraph of $G^{S.C}_{\text{cud}}$ induced by nodes v has only full lines and a missing (i,j)-full-line if and only if the generating process implies for $i,j \in v$

$$i \perp\!\!\!\perp j \mid C \cup v \setminus \{i,j\},$$

i.e. it is the conditional concentration graph of Y_v given Y_C.

Node pairs of offspring u may no longer have a simple edge, but may have a *composite edge* consisting of two components, namely of an arrow and of a dashed line. Whenever an edge is composite, its components are to be thought of as different paths. There is an (i,j)-dashed line missing in the subgraph of $G^{S.C}_{\text{cud}}$ induced by nodes u if and only if the joint conditional distributions of Y_i and Y_j which factorized in (1) still factorizes after marginalising over M and conditioning on C.

In the part of $G^{S.C}_{\text{cud}}$ not involving ancestors v there is an (i,j)-arrow missing if and only if j is not a M-line ancestor of i in the generating graph and no ancestor-like relation is generated for them by conditioning on C and marginalising over M. As we shall see no conditional independence statement for pair (i,j) needs to hold even when (i,j)-edge components of both type are missing in $G^{S.C}_{\text{cud}}$.

Finally, the *graph of projecting* each component of Y_S on Y_C is a fully directed graph without edges for nodes within S and within C. The (i,j)-arrow is missing in it if and only if the generating process implies for $i \in S$ and $j \in C$

$$i \perp\!\!\!\perp j \mid C \setminus j.$$

As discussed in Sections 3 and 4 in the special case when S consists of offspring H and ancestors D of C outside C, i.e.if $S = V \setminus C$, the linear system (4) implies $\Pi_{H|D.C} = -A^{-1}_{HH}A_{HD}$, so that there is a structural zero in the graph of projecting $Y_{H|C}$ on $Y_{D|C}$ if and only if j is a not a H-line ancestor of i.

This example is well suited to illustrate the type of reasoning needed for proving that structural zeros in matrices for linear Gaussian systems imply missing edges in corresponding graphs of arbitrary distributions generated over the same parent graph of A. Thus, for the graph of projecting $Y_{H|C}$ on $Y_{D|C}$ it is to be proven that the factorization (1) implies $i \perp\!\!\!\perp j \mid D \cup C \setminus j$ if node j is not an H-line ancestor of i.

More precisely, for $i \in H$, the offspring set of C, and $j \in D$, the ancestor set of C outside C, we have

$$f_{i|C,D} = \int f_{i|C,D,\mathcal{P}} f_{\mathcal{P}|C,D} d\mathcal{P},$$

where P denotes the parents of node i not in $L = C \cup D$, but in $\{i+1, \ldots, d_H\}$. The first factor on the right-hand side cannot depend on j, for otherwise node j would be a parent of i. To deal with the second factor we integrate in the order of the generating process, i.e. we start with $l = d_H$. The density of each component $l \in \{i+1, \ldots, d_H\}$ depends only on $\mathrm{par}(l)$, the parents of this component, taken conditionally on $\{l+1, \ldots, d_H\}$ and $C \cup D$. Now none of these densities can depend on j, for otherwise j would be an H-line ancestor of i in the parent graph. Thus, $f_{i|C,D}$ does not depend on j, as was to be proved.

Thus, a matrix condition assures for all Gaussian distribution that $i \perp\!\!\!\perp j \mid D \cup C \setminus j$ is implied by the generating process. The condition reformulated for graphs implies the same independency for general distributions, provided they are generated over the same parent graph. This is set out in detail for various other graphs in the following Sections.

5.2 Missing edges in the overall ancestor graph, G^V_{anc}

The following statements are equivalent:

(i) There is a structural (i, j)-zero in the inverse B of the triangular matrix A.

(ii) The (i, j)-arrow is missing in the ancestor graph of A.

(iii) In the parent graph of A there is no direction-preserving path pointing to node i from node j.

This implies that if there is an (i, j)-structural zero in the matrix B then for arbitrary distributions generated over the graph $i \perp\!\!\!\perp j \mid \{j+1, \ldots, d_V\}$.

The joint conditional density of all variables of interest here is

$$f_{ij|j+1,\ldots,d_V} = f_{i|j,j+1,\ldots,d_V} f_{j|j+1,\ldots d_V}. \tag{23}$$

Now in general not all the parents of i are in the conditioning set for the first factor. Therefore we partition the parents of i into those in $\mathcal{P} = \{i+1, \ldots, j-1\}$ and the remainder in $\{j, j+1, \ldots, d_V\}$. Then the first factor is

$$f_{i|j,j+1,\ldots,d_V} = \int f_{i|\mathcal{P},j,\ldots,d_V} f_{\mathcal{P}|j,\ldots,d_V} d\mathcal{P}. \tag{24}$$

Now neither factor on the right-hand side of (24) can involve j. For otherwise by *(iii)* j would be either a parent or an ancestor of i. That is (24) is a function of i alone and, because the second term on the right-hand side of (23) does not not involve i, the expression (23) factorizes into a function of i times a function of j proving the required independency.

For the claimed equivalences *(i)* to *(iii)* note that in Section 2 the inverse of a $r \times r$ triangular matrix A was obtained by sweeping A on all its rows and columns and the necessary and sufficient condition for having a structural zero in $B = A^{-1}$ was illustrated for pair (1,4). More generally, the condition is that there is no ordered sequence $(i = i_0, i_1, \ldots, i_t, i_{t+1} = j)$ such that each adjacent pair is a free parameter in A. Statements *(ii)* and *(iii)* are equivalent by definition for linear systems.

For general systems they are also equivalent since $\mathcal{B}$, the edge matrix of the ancestor graph G^V_{anc}, can be obtained directly from $\mathcal{A}$, the edge matrix of the parent graph G^V_{par} as follows. Every transition-oriented V-configuration in the parent graph is completed by an arrow, until no such V-configuration remains. The arrow is inserted so that it shortens the path via the transition-node, i.e. so that it keeps the same direction as the two arrows in the V-configuration that is completed. The matrices in (6) illustrate the result. To see the connection of inverting A to completing transition-oriented V-configurations note that the edge matrix $\mathcal{B}$ can be obtained from the edge matrix $\mathcal{A}$ completing off-diagonal submatrices (i, j, k) by inserting (i, k)-ones as long as an (i, k)-zero coincides with ones in positions (i, j) and (j, k).

As is set out in the next two Sections both graphs G^V_{anc} and G^V_{par} determine in a dual way which edges are missing in the overall covariance graph G^V_{cov} and in the overall concentration graph G^V_{con}, respectively.

5.3 Missing edges in the overall concentration graph, G^V_{con}

The following statements are equivalent:

(i) There is a structural (i,j)-zero in the inner product of the triangular matrix A, i.e. in $A^T A$.

(ii) The (i,j)-full-line is missing in the overall concentration graph G^V_{con} of A.

(iii) In the parent graph of A the node pair (i,j) has no edge and no common collision node.

This implies that if for all possible Gaussian distributions generated over the parent graph of A the (i,j)-element of the concentration matrix vanishes, then for all distributions generated over the same graph $i \perp\!\!\!\perp j \mid \{1,\ldots,d_V\} \setminus \{i,j\}$.

The generating process uses $f_{i|i+1,\ldots,d_V} = f_{i|\text{par}(i)}$ so that the (i,j)-edge missing by (iii) in the parent graph means $i \perp\!\!\!\perp j \mid c$ with $c = \{i+1,\ldots,d_V\} \setminus j$ and $f_{i|j,c} = f_{i|c}$. Now also by (iii), for all $h < i$ at most one of i and j is in the parent set of h. For otherwise (iii) would be contradicted. Hence $h \perp\!\!\!\perp i \mid (j,c)$ or $h \perp\!\!\!\perp j \mid (i,c)$ and $f_{h|i,j,c} = f_{h|j,c}$ or $f_{h|i,j,c} = f_{h|i,c}$.

We now argue one step at a time to extend the conditioning set c for i by adding in sequence h of $(i-1,\ldots,1)$. For example with $i \perp\!\!\!\perp j \mid c$ and $h \perp\!\!\!\perp i \mid (j,c)$ we have that

$$f_{h,i,j|c} = f_{h|i,j,c} f_{i|j,c} f_{j|c} = f_{h|j,c} f_{i|c} f_{j|c} = f_{h,j|c} f_{i|c},$$

i.e. $(h,j) \perp\!\!\!\perp i \mid c$. Therefore,

$$f_{i,j|h,c} = f_{h,i,j|c}/f_{h|c} = f_{h,j|c} f_{i|c}/f_{h|c} = f_{j|h,c} f_{i|c},$$

so that $i \perp\!\!\!\perp j \mid (h,c)$. By the same type of reasoning, we have from $i \perp\!\!\!\perp j \mid c$ and $h \perp\!\!\!\perp j \mid (i,c)$ that $(h,i) \perp\!\!\!\perp j \mid c$ and hence also $i \perp\!\!\!\perp j \mid (h,c)$. After all $i-1$ steps of enlarging the conditioning set of i the required independency $i \perp\!\!\!\perp j \mid \{1,\ldots,d_V\} \setminus \{i,j\}$ follows. A slightly more general result is given in Appendix 2 to illustrate the appealing reasoning in terms of graphs instead of densities.

For the claimed equivalences (i) to (iii) note that the triangular decomposition (5) of the concentration matrix in a linear system (4) gives $\text{con}_V(Y) = \Sigma^{-1} =$

$A^T\Delta^{-1}A$. Since Δ is a diagonal matrix its inverse does not affect structural zeros in this matrix product.

The explicit form of the matrix product A^TA shows that a structural (i, j)-zero in A remains a structural zero in A^TA if and only if there is no index $h < i$ such that a_{hi} and a_{hj} are both free parameters in A. This necessary and sufficient condition for obtaining an (i, j)-structural zero in the inner product of A is for linear systems equivalent to statements (ii), (iii) by definition.

For general systems they are also equivalent since the overall concentration graph can be obtained directly from G^V_{par} with edge matrix $\mathcal{A}$ by first completing every collision-oriented V-configuration by an edge and then replacing all edges by full lines.

Since a concentration matrix is symmetric, its indicator matrix for structural zeros can again be stored in an upper triangular matrix. The edge matrix of G^V_{con}, is defined here as an upper triangular matrix of zeros and ones with an (i, j)-zero if and only if the (i, j)-edge is missing in G^V_{con}. It coincides by definition with the upper triangular part of the indicator matrix of structural zeros in A^TA. To see the connection to completing collision-oriented V-configurations, note that the edge matrix of the concentration graph can be obtained from the edge matrix $\mathcal{A}$ completing off-diagonal submatrices (h, i, j) by inserting (i, j)-ones as long as an (i, j)-zero coincides with ones in positions (h, i) and (h, j).

5.4 Missing edges in the overall covariance graph, G^V_{cov}

The following statements are equivalent:

(i) There is a structural (i, j)-zero in the outer product of the triangular matrix $B = A^{-1}$, i.e. in BB^T.

(ii) The (i, j)-dashed line is missing in the overall covariance graph G^V_{cov} of A.

(iii) In the ancestor graph of A the node pair (i, j) has no edge and no common source node.

(iv) In the parent graph of A there is no collisionless path between nodes i and j.

This implies that if for all possible Gaussian distributions generated over the parent graph of A the (i, j)-element in the overall covariance matrix vanishes, then for all distributions generated over the same graph $i \perp\!\!\!\perp j$.

To see this note first that the absence of a collisionless path between nodes i and j in (iv) is equivalent to the sets ($i \cup$ ancestors of i) and ($j \cup$ ancestors of j) being disjoint. Assume first that there is no collisionless path between i and j. Now, the distribution of i is recursively formed via (1) by contributions of ancestors of i. If j or one of its ancestors were an ancestor of i then there would be a collisionless path between i and j contradicting the assumption. Similarly, the distribution of j is generated from its ancestors. Since j is a potential ancestor of i, the node i cannot be in the ancestor set of j. If one of the ancestors of i were also an ancestor of j or of one of the ancestors of j then there would at the same time be a common-source path between i and j contradicting again the assumption. Conversely if the two sets formed by each node with its ancestors are disjoint then every possible collisonless path would generate an overlapping set of ancestors for i and j, again a contradiction to the assumption.

The disjointness implies with (1) that the joint marginal density of the two sets of variables factorizes so that ($i \cup$ ancestors of i)$\perp\!\!\!\perp$($j \cup$ ancestors of j). After marginalising over the ancestors of i and j we have the required marginal independence $i \perp\!\!\!\perp j$.

For the claimed equivalences (i) to (iv) note that the triangular decomposition (5) of the covariance matrix of a linear recursive system (4) gives $\text{cov}(Y) = \Sigma = B\Delta B^T$. Since Δ is a diagonal matrix it does not affect structural zeros in this matrix product.

The explicit form of the matrix product BB^T shows that a structural (i, j)-zero in B remains a structural zero in BB^T if and only if there is no index $k > j$ such that b_{ik} and b_{jk} are both free parameters in B. This necessary and sufficient condition for obtaining an (i, j)-structural zero in the outer product of B is for linear systems equivalent to statements (ii), (iii) by definition. The equivalence to statement (iv) follows from a combination of results here with those of Section 5.1.

For general systems statements (i) to (iv) are also equivalent since the overall covariance graph can be obtained directly from G^V_{anc} with edge matrix $\mathcal{B}$ by first completing every source-oriented V-configuration by an edge, then replacing all edges by dashed lines.

Since a covariance matrix is symmetric, its indicator matrix for structural zeros can again be stored in an upper triangular matrix. The edge matrix of G^V_{cov}, is defined here as an upper triangular matrix of zeros and ones with an (i, j)-zero if and only if the (i, j)-edge is missing in G^V_{cov}. By definition it coincides with the upper triangular part of the indicator matrix of structural zeros in BB^T. To see the connection to completing source-oriented V-configurations note that the edge matrix of the covariance graph can be obtained from the edge matrix $\mathcal{B}$ completing off-diagonal submatrices (i, j, k) by inserting (i, j)-ones as long as an (i, j)-zero coincides with ones in positions (i, k) and (j, k).

5.5 Missing edges in q-line ancestor graphs

The following statements are equivalent:

(i) There is a structural (i, j)-zero in a triangular matrix A^* obtained with the swt-operator (1) by sweeping A on rows and columns q.

(ii) The (i, j)-edge is missing in the q-line ancestor graph of A.

(iii) In the parent graph of A there is no (i, j)-edge and no direction-preserving path between nodes i and j having all nodes along it in q.

This implies that if for all possible Gaussian distributions generated over the parent graph of A the (i, j)-element of the matrix A after sweeping on rows and columns of M is no longer a structural zero, then for all distributions generated over the same graph the conditional independence of a potential ancestor j of i present in the generating process is no longer necessarily preserved after having reduced the conditioning set by marginalising over $i, \dots, j-1$.

For the claimed equivalences note that in Section 4.2 details were given for A_{LL} swept on an arbitrary subset $q = L \setminus K$, where $A_{LL}Y_L = \varepsilon_L$ is a subsystem of linear equations (4) with uncorrelated residuals obtained by marginalising over offspring

H of L, i.e. over adjacent nodes H in $V = (H, L)$. Equations (15) to (18) contain the explicit forms of

$$(A_{LL} \text{ swt } q) = \begin{pmatrix} (A_{LL} \text{ swt } q)_{K,K} & (A_{LL} \text{ swt } q)_{K,q} \\ (A_{LL} \text{ swt } q)_{q,K} & (A_{LL} \text{ swt } q)_{q,q} \end{pmatrix} = \begin{pmatrix} A_{KK.q} & A_{Kq}A_{qq}^{-1} \\ -A_{qq}^{-1}A_{qK} & A_{qq}^{-1} \end{pmatrix}.$$

These components look quite different, since they are a sum of matrices, matrix products, and an inverse. However, regarding their interpretation in terms of the edge matrix, $\mathcal{A}$, of the parent graph they only differ with respect to the location of an (i, j)-element. As mentioned before node j is said to be a q-line ancestor of i if there is an (i, j)-edge in the parent graph or a direction-preserving path with all nodes along it in q. Thus there is a structural (i, j)-zero in $A_{KK.q}$ if and only if for $j \in K$ is not a q-line ancestor of $i \in K$. Similarly, there is is a structural (i, j)-zero in $\Theta_{K|q} = A_{Kq}A_{qq}^{-1}$ if and only if $j \in q$ is not a q-line ancestor of $i \in K$, and so on.

For general systems statements (i) to (iii) are also equivalent since the q-line ancestor graph, $G^L_{\text{anc}(q)}$, can be obtained directly from the parent graph with edge matrix $\mathcal{A}$ by completing every transition-oriented V-configuration with common neighbour within q until none is left. Since $G^L_{\text{anc}(q)}$ is a directed acyclic graph its edge matrix is just as defined previously for directed acyclic graphs. It coincides by definition with the indicator matrix of structural zeros in the upper triangular matrix $(A \text{ swt } q)$. The q-line ancestor graph can be regarded as an intermediate step for deriving conditional independence statements, in the same way as the following graphs of connected univariate dependences.

5.6 Missing edges in the graph of connected univariate dependencies G^K_{cud} for the marginal distribution of Y_K

The general definition in Section 5.1 of edges in a graph of connected univariate dependencies contains the graph for a marginal distribution as a special case, i.e. if $C = \emptyset$. In particular it implies for i and j both in K that an (i, j)-arrow is missing in the graph of connected univariate dependencies, G^K_{cud} of A, if and only if it is missing in the subgraph induced by nodes K in the q-line ancestor graph of A.

The following statements are also equivalent for i and j both in K:

(i) There is a structural (i,j)-zero in the outer product of $\Theta_{K|q} = (A \text{ swt } q)_{K,q}$.

(ii) An (i,j)-dashed-line is missing in the graph of connected univariate dependencies G^K_{cud} of A.

(iii) In the q-line ancestor graph of A the node pair (i,j) has no common source node in q.

(iv) In the parent graph of A node pair i,j has no common source path with all nodes along it in q.

For the claimed equivalences of (i) to (iv) note that the marginal joint distribution for variables of an arbitrary subset K of L is for linear systems given by the second equation in (13). The parameters in the system are $A_{KK.q}$ and Δ_{KK-q}, since

$$A_{KK.q}Y_K = \varepsilon_{K-q} = \varepsilon_K - \Theta_{K|q}\varepsilon_q, \quad \text{cov}(\varepsilon_{K-q}) = \Delta_{KK-q} = \Delta_{KK} + \Theta_{K|q}\Delta_{qq}\Theta^T_{K|q}.$$

The interpretation of a swept matrix given in the previous section shows that $A_{KK.q}$ keeps track of edges present in the parent graph and of direction-preserving paths with all nodes along it in q while Δ_{KK-q} keeps tracks of common-source paths with all nodes along it in q.

The equivalence of the statements follows for linear and for general systems by the same type of reasoning given in the previous sections.

5.7 Missing edges in the covariance and concentration graph of Y_K

To complete the discussion of the marginal distribution of Y_K we note that, as for any subset of nodes, the covariance graph, G^K_{cov}, is the subgraph induced by nodes K in the overall covariance graph.

For the marginal concentration graph, G^K_{con}, the equivalence of the following statements is helpful, where as before $L = K \cup q = V \setminus$ offspring of L.

(i) There is a structural (i,j)-zero in the matrix product $A^T_{KK.q}\Delta^{-1}_{KK-q}A_{KK.q}$.

(ii) An (i,j)-full-line is missing in the marginal concentration graph of Y_K, in G^K_{con}.

(iii) In G^K_{cud} node pair (i,j) is not connected by a pure collision path.

(iv) In G^L_{con} there is no (i,j)-edge and there is no path between nodes i and j with all nodes along it in q.

(v) In the subgraph induced by L in the parent graph of A the pair (i,j) has no edge, no common collision node, and no path connecting i and j with every transition node and every source node in q. (Hence, every collision node along the path is in L.)

Statement (i) is a necessary and sufficient condition that $i \perp\!\!\!\perp j \mid K \setminus (i,j)$ for all Gaussian distributions generated over the parent graph of A. Marginalising over the offspring H of $L = K \cup q$ does not affect the remaining recursive system in nodes L. Thus, we need to show that the equivalence with the path conditions implies $i \perp\!\!\!\perp j \mid L \setminus \{i,j,q\}$ for arbitrary distributions genererated over the same parent graph.

To see this we have first by condition (v) and the results of Section 5.3 that the absence of an edge and a common collision node in L for pair (i,j) implies $i \perp\!\!\!\perp j \mid L \setminus \{i,j\}$ for arbitrary distributions. This is a property of the overall concentration graph to the subgraph induced by nodes L in the parent graph which is directed acyclic.

To verify the required independency we have to remove the nodes in q from the conditioning set, i.e. to obtain the concentration graph of nodes $L \setminus q$. To examine marginalising over q we argue by mathematical induction on the number d_q, say, of the nodes in q. From (v) we consider only paths in the subgraph induced by nodes L in the parent graph and we call a path between any two nodes i and j active if every transition node and every source node along it is in q and every collision node is in L. We take as the induction hypothesis that the absence of an active path between i and j is a sufficient condition for the required independence.

If $d_q = 1$ so that q consists of a single node γ then by the result in Appendix 2 absence of an active path is a sufficient condition for retaining the conditional independence after removing γ from the conditioning set. The reasoning by induction proceeds by considering first active paths within the $d_q - 1$ nodes to the union of i, j, then if none of the resulting paths is active between i and j, the three node configuration between i and j and the last node of q is considered to complete the argument.

For the claimed equivalences of (i) to (v) note that in Section 4.2 it was shown that there is a free parameter in Δ^{-1}_{KK-q} if and only if there is a sequence of indices connecting i and j which corresponds to free parameters in Δ_{KK-q}. In the graph G^K_{cud} this corresponds to the presence of a path of only dashed lines between i and j. By conditioning on nodes K, i.e by inverting Δ_{KK-q}, every V-configuration of dashed lines in G^K_{cud} is closed until none are left. Since the concentration graph of Y_K is $\Sigma^{-1}_{KK} = A^T_{KK.q}\Delta^{-1}_{KK-q}A_{KK.q}$ it has a structural zero if and only if i and j are not connected by a path in Δ_{KK-q} and not by an arrow pointing from i or from j or from both to a path in Δ_{KK-q}. This enumerates the four possible types of pure collision path in G^K_{cud}.

Similarly, it was explained in Section 4.2 why by marginalising in Σ^{-1}_{LL} over q a structural (i,j)-zero is preserved if and only if there is no sequence of indices in q connecting i and j which corresponds to free parameters in Σ^{-1}_{LL}. In the graph G^L_{con} this corresponds to the absence of an edge and of a path with all nodes along it in q. By marginalising over nodes q the V-configurations of full lines in G^L_{con} with common neighbour in q are closed until none are left. This provides two different routes of constructing the concentration graph of Y_K.

The conditions in (v) are just a reformulation for the parent graph to assure that after marginalising over q no pure collision path is generated in G^K_{cud} and that after conditioning on C there is no collisionless path between i and j in L wholly outside C. Finally we note that by the definitions of graphs given in Section 5.1 a missing (i,j)-edge in the concentration graph of Y_K is equivalent to a missing (i,j)-edge in the conditional covariance graph of $H \cup \{i,j\}$ given $C' = K \setminus \{i,j\}$, i.e. as aspecial case of the more general conditional independence statement discussed in Section 5.9 below.

5.8 Missing edges in the graph of connected univariate dependencies $G^{u|K}_{\text{cud}}$ for the conditional distribution of Y_u given Y_K

The general definition in Section 5.1 of edges in a graph of connected univariate dependencies implies for i and j both in u that there is a structural (i, j)-zero in the triangular matrix $A_{uu.p} = (A \text{ swt } M)_{u,u}$ and an (i, j)-arrow is missing in the graph of connected univariate dependencies, $G^{u|K}_{\text{cud}}$ of A if and only if it is missing in the subgraph induced by nodes u in the M-line ancestor graph of A.

The following statements are equivalent for i in u and for j in K:

(i) There is a structural (i, j)-zero in the matrix product -$A_{uu.p}\Pi_{u|K} = A_{uK.M} - \Delta_{uK-M}\Delta^{-1}_{KK-M}A_{KK.M}$.

(ii) An (i, j)-arrow is missing in the graph of connected univariate dependencies, $G^{u|K}_{\text{cud}}$ of A.

(iii) In the graph of marginal connected univariate dependences, $G^{V\setminus M}_{\text{cud}}$ of A there is no (i, j)-edge-component and no node i' in K having an (i', j)-arrow and connecting to i via a path of dashed-lines with all nodes along it in K.

For both nodes i and j in u the following statements are equivalent:

(i) There is a structural (i, j)-zero in the matrix product $A_{uu.p}\Sigma_{uu.K}A^T_{uu.p} = \Delta_{uu-M} - \Delta_{uK-M}\Delta^{-1}_{KK-M}\Delta^T_{uK-M}$.

(ii) An (i, j)-dashed-line is missing in the graph of connected univariate dependencies $G^{u|K}_{\text{cud}}$ of A.

(iv) In the graph of marginal connected univariate dependences, $G^{V\setminus M}_{\text{cud}}$ of A there is no (i, j)-dashed-line and no path of dashed-lines connecting i and j with all nodes along it in K.

For the claimed equivalences some key results in Sections 4.4 and 4.5 are that marginalising over M is for offspring of C the same as marginalising just over $M_{\text{off}} = p$ and for ancestors of C the same as marginalising just over $M_{\text{anc}} = q$.

The explicit expressions for components of the matrix A swept on rows and columns M and of $\text{cov}(\varepsilon_{V\setminus M-M})$ permit the following interpretation of edges present in $G^{u|K}_{\text{cud}}$.

Every (i, j)-dashed line is in linear systems a covariance of residuals. In general systems it corresponds to a connection between the two univariate conditional distributions of Y_i and Y_j induced by marginalising over all nodes along a common-source path between i and j present in the parent graph of A or over a common-source like path generated by conditioning nodes of K, i.e. over a new collisionless path between arrows pointing to i and j in the parent graph of A.

Every additional (i,j)-arrow corresponds in general systems to a connection between the two univariate conditional distributions of Y_i and Y_j which is induced by marginalising over all nodes along a descendant-ancestor paths present between i and j in the parent graph of A or over all nodes along a descendant-ancestor like path generated by conditioning on K, i.e. a new collisionless path between an arrow pointing at i in the parent graph of A and starting from j in K.

5.9 Missing edges in graphs for the conditional distribution of Y_S given Y_C

The connected univariate dependence graph of $Y_{u|C}$ given $Y_{v|C}$ and the conditional concentration graph of $Y_{v|C}$ is the subgraph induced by nodes $S = u \cup v$ in the graph $G^{u|K}_{\text{cud}}$ combined with the marginal concentration graph G^{K}_{con}. Keeping $A_{uu.p}$ and taking submatrices of $\Pi_{u|K}$ and of Σ^{-1}_{KK} for $v = K \setminus C$ in linear systems (Section 4.4) corresponds in general systems to keeping subgraphs induced by nodes within S in the graph of $G^{u|K}_{\text{cud}}$ combined with G^{K}_{con}. Keeping $A_{uu.p}$ and taking submatrices of. For example for $u = \{1, 3, 4\}$ and $v = \{7, 10\}$ in Figure 1, it is the subgraph induced by $S = u \cup v$ in Figure 4.

For the conditional concentration graph, $G^{S.C}_{\text{con}}$, and the covariance graph, $G^{S.C}_{\text{cov}}$, matrix results for linear systems in Section 4.6 are again translated into modifying graphs and edge matrices. No new type of reasoning is needed. We give here just a general summary of important aspects and an algorithmic matrix formulation for deriving the edge matrix of $G^{S.C}_{\text{cov}}$ directly from the edge matrix $\mathcal{A}$ of the parent graph.

The graph $G^{S.C}_{\text{con}}$ is the subgraph induced by nodes S in the marginal concentration matrix of $G^{V \setminus M}_{\text{con}}$. The graph $G^{S.C}_{\text{cov}}$ is the subgraph induced by nodes S in the conditional covariance matrix of $G^{(V \setminus C).C}_{\text{cov}}$. Their construction from the parent graph requires a different treatment of certain nodes: for $G^{S.C}_{\text{con}}$ offspring and ancestors of $S \cup C$ are needed while for $G^{S.C}_{\text{con}}$ nodes which are offspring and which are ancestors of C are to be distinguished.

When the conditional covariance graph of ancestors outside the conditioning set is constructed from their concentration graph every path present in the conditional concentration graph gets closed, i.e. every V-configuration is completed, and no edges are removed since the conditioning is not undone. The resulting conditional covariance graph of the ancestors consists exclusively of complete, nonoverlapping subgraphs. In the matrices of linear systems this corresponds to complete block-diagonal structure. In general it is not possible to reverse this construction step, since the possible additional missing edges in a conditional concentration graph of ancestors cannot be recovered from their conditional covariance graph.

Residuals obtained after marginalising alone, like ε_{u-M}, get correlated by marginalising over all ancestor nodes along a common source path unless there is parametric cancellation. Residuals like η_u, obtained by marginalising and conditioning, are independent of residuals of potential ancestors of u. They may not yet get correlated after marginalising alone but only after there is in addition conditioning on nodes which are common descendants of ancestors of u. But nodes of this conditioning set are not in u and they are not descendants of any components of u, otherwise there would be a path from u into the conditioning set contradicting the definition of u as offspring of the conditioning set C. Therefore, by conditioning on u such correlations cannot be undone and no new correlations can get induced among potential ancestors of u. Thus, the concentration graph of the residuals can be constructed from their covariance graph by completing every V-configuration until none is left. The resulting concentration graph of the residuals consists exclusively of complete, nonoverlapping subgraphs. Which additional edges are missing in the covariance graph of the residuals cannot be recovered from their concentration graph, i.e. the construction step cannot be reversed.

From the graphs of connected univariate dependences for the conditional distribution of Y_u given Y_C the concentration graph of Y_S given Y_C and the covariance graph of Y_S given Y_C, can both be obtained directly. For this the matrix formulations for linear systems, given in Section 4.6, just need to be reformulated in terms of graphs. But in general it is not possible to construct the concentration graph of Y_S given Y_C directly from its covariance graph or the covariance graph of Y_S given Y_C directly from its concentration graph, since edges may get added as well as removed.

To prove the equivalence of a path condition on the parent graph for a structural zero in the conditional covariance graph of Y_S given Y_C to matrix conditions for a structural zero in $\Sigma_{(V\setminus C,V\setminus C).C}$ the overall node set is partitioned into H, C, D, where H are offspring of the conditioning set C and D are ancestors of C outside C.

The path condition given by Wermuth and Cox (1998) says that there is an edge missing in $G^{S.C}_{\text{cov}}$ if and only if in the parent graph G^{V}_{par} modified by conditioning on C there is no collisionless path wholly outside C. To modify by conditioning on C means that in the subgraph induced by the ancestors D of C in the parent graph every missing (i,j)-edge is joined by a full line provided it has a common collision node in D or in C, i.e. outside the offspring H of C. To modify by conditioning and then changing all resulting edges within D to full lines is the step of deriving G^{D}_{con} of A, the concentration graph of Y_D from the parent graph of A to G^{V}_{par}. In the linear case this is the modification of the generating system (4) to obtain (22).

Starting from system (22) we give first conditions for structural zeros for a multivariate regression chain of $Y_{H|C}$ regressed on $Y_{D|C}$ and derive next from it structural zeros in the covariance matrix of $Y_{V\setminus C}$ given Y_C.

From (22) we get

$$Y_{H|C} + A^{-1}_{HH}A_{HD}y_{D|C} = A^{-1}_{HH}\varepsilon_H, \quad Y_{D|C} = \eta_D. \tag{25}$$

and the following statements as equivalent for $i \in H$ and $j \in D$.

(*i*) There is a structural (i,j)-zero in the matrix product $\Pi_{H|D.C} = -A^{-1}_{HH}A_{HD}$.

(*ii*) An (i,j)-arrow is missing in the graph of projecting $Y_{H|C}$ on $Y_{D|C}$ as implied by the parent graph of A.

(*iii*) In the parent graph of A node j is not a parent of node i nor a H-line ancestor of i.

The following statements are equivalent for $i, j \in H$.

(i) There is a structural (i,j)-zero in the matrix product $A_{HH}^{-1}A_{HH}^{-T}$.

(ii) An (i,j)-dashed line is missing in $G_{\text{cov}}^{H.D,C}$ of A.

(iii) In the H-line ancestor graph, $G_{\text{anc(H)}}^{V}$ of A, node pair (i,j) has no (i,j)-edge and no common source node.

Together with the concentration graph of $Y_{D|C}$ this gives the graphical representation of structural zeros in

$$(\Sigma \text{ swp } C, D)_{L,L} = \begin{pmatrix} \Sigma_{HH.DC} & \Pi_{H|D.C} \\ . & -\Sigma_{DD.C}^{-1} \end{pmatrix},$$

with dashed lines for the covariance graph within H, with arrows for the matrix of regression coefficients between H and D and with full lines within D, i.e. a joint response chain graph as in Figure 2 of only simple edges. It consists of three parts, of $G_{\text{cov}}^{H.C,D}$ of A, of $G_{\text{con}}^{D.C}$ of A, and of the graph of projecting $Y_{H|C}$ on $Y_{D|C}$ as implied by system (1). The proofs in Section 5.4 for an overall covariance graph, in Section 5.3 for an overall concentration graph, and at the end of Section 5.1 for projecting apply with small modifications to show that each structural zero in matrices of the joint distribution in (25) mean corresponding conditional independence statements for arbitrary distributions generated over the same parent graph.

Now, the covariance matrix of $Y_{V\setminus C}$ given Y_C is obtained from $(\Sigma \text{ swp } C, D)_{L,L}$ by resweeping on rows and columns D so that a structural zero is implied and a corresponding (i,j)-dashed line is missing in $G_{\text{cov}}^{(V\setminus C).C}$ of A if and only if in the joint response chain graph of (25) there is for any pair (i,j) no path between i and j with all nodes along it in D and there is

(i) for $i, j \in D$ no (i,j)-full-line;

(ii) for $i \in H$ and $j \in D \cup H$ no (i,j)-arrow;

(iii) for $i, j \in H$ no (i,j)-dashed-line.

Every path with all nodes along it in D is a collisionless path since every edge in a concentration graph is a full line and since edges between D and H are arrows pointing into H. Every single edge in (i) to (iii) is defined to be a collisonless path since it is a path of length one, irrespective of the type. And, finding the H-line ancestors to get from the system of univariate connected dependencies (22) to the multivariate regression in (25) means to shorten direction-preserving paths with all nodes along it in H into an arrow and to close every common source path with all nodes along it in H by a dashed line. Thus, every edge present in $G_{\text{cov}}^{(V\setminus C).C}$ is generated by a collisionless path in $G_{\text{cud}}^{(V\setminus C).C}$ of A, i.e. by a collisionless path present in the parent graph before conditioning or generated after having conditioned on C. Every remaining path between i and j in the parent graph is a collision path via offspring of C. This proves the equivalence of the path criterion to constructing $\Sigma_{(V\setminus C,V\setminus C).C}$ by matrix transformations.

As shown at the end of Section 4.6 the structural zeros in $\Sigma_{(V\setminus C,V\setminus C).C}$ are those in the outer product of the upper triangular matrix

$$F = \begin{pmatrix} A_{HH}^{-1} & -A_{HH}^{-1}A_{HD}F_{DD} \\ 0 & F_{DD} \end{pmatrix} = \begin{pmatrix} A_{HH}^{-1} & -A_{HH}^{-1}A_{HD} \\ 0 & I_{DD} \end{pmatrix} \begin{pmatrix} I_{HH} & 0 \\ 0 & F_{DD} \end{pmatrix},$$

since $\Sigma_{(V\setminus C,V\setminus C).C} = F\Delta^+F^T$ and Δ^+ is diagonal. And, because the indicator matrix $\mathcal{F}$ of structural zeros in F is also the edge matrix of the overall ancestor graph in the relevant coverering system $EY_{(V\setminus C).C} = \varepsilon^*$ our results in Section 5.4 can be applied to the overall covariance graph of this this system which has identical edge matrix as the covariance graph $G_{\text{cov}}^{(V\setminus C).C}$ of A. Thus, $Y_i \perp\!\!\!\perp Y_j \mid Y_C$ is implied for general distributions if and only if there is a zero in $\mathcal{F}^+$, the indicator matrix of structural zeros in FF^T, and, equivalently, when the above path criterion for the generating directed acyclic graph in Y is not satisfied.

5.10 Obtaining the edge matrix of $G_{\text{cov}}^{(V\setminus C).C}$ directly from the one of G_{par}^V

The edge-matrix $\mathcal{F}^+$ of the conditional covariance graph of Y_S given Y_C can be constructed directly from the edge-matrix $\mathcal{A}$ of the parent graph by using these results. The modifications of the edge matrix $\mathcal{A}$ involve nothing but completing different types of V-configuration, i.e. replacing a zero by a one in off-diagonal submatrices of three nodes with two ones and one zero. The construction steps are as follows, where as before $V = H \cup C \cup D$, $L = C \cup D$, $V \setminus C = H \cup D$ and nodes (h, i, j, k) are taken to be in increasing order.

(i) Remove within the (D, D)-submatrix of $\mathcal{A}_{LL}$ every (i, j)-zero which has for h in L a one in positions (h, i) and (h, j).

(ii) Remove in the matrix obtained in step (i), every V-configuration until none is left. Call the resulting matrix $\mathcal{F}_{DD}$.

(iii) Remove in $\mathcal{A}_{HH}$ every transition-oriented V-configuration until none is left. Call the resulting matrix $\mathcal{F}_{HH}$.

(iv) Remove in $\mathcal{A}_{HD}$ every transition-oriented V-configuration with $\mathcal{F}_{HH}$ and $\mathcal{F}_{DD}$ until none is left. Call the resulting matrix $\mathcal{F}_{HD}$ and call $\mathcal{F}$ the upper triangular matrix consisting of $\mathcal{F}_{HH}$, $\mathcal{F}_{HD}$ and $\mathcal{F}_{DD}$.

(v) Remove in $\mathcal{F}$ every source-oriented V-configuration.

The indicator matrix $\mathcal{F}$ gives the structural zeros in the overall ancestor graph for $Y_{(V\setminus C)|C}$. The modified indicator matrix $\mathcal{F}$ is the edge matrix $\mathcal{F}^+$ of $G_{\text{cov}}^{(V\setminus C).C}$, and the edge matrix of $G_{\text{cov}}^{S.C}$ is the submatrix $\mathcal{F}^+_{\mathcal{SS}}$.

To understand the construction note that with (i) the edge matrix of $G_{\text{con}}^{D.C}$ is obtained or, equivalently, the indicator matrix of structural zeros in $\Sigma_{DD.C}^{-1} = (A_{LL}^T \Delta_{LL} A_{LL})_{D,D}$ is formed. With (ii) the edge matrix $\mathcal{F}_{\mathcal{DD}}$ of $G_{\text{cov}}^{D.C}$ results. Because in this step every path present in $G_{\text{con}}^{D.C}$ is closed the resulting covariance graph consists of nonoverlapping complete subgraphs. Therefore $\mathcal{F}_{\mathcal{DD}}$ represents at the same time the structural zeros in the triangular decomposition $\Sigma_{DD.C} = F_{DD}\Delta^*_{DD}F^T_{DD}$. With ($iii$) the edge matrix $\mathcal{F}_{HH}$ of the overall ancestor graph of $Y_{H|L}$ is obtained, or, equivalently, the structural zeros in $F_{HH} = A_{HH}^{-1}$. With (iv) the indicator matrix $\mathcal{F}_{HD}$ of structural zeros in $F_{HD} = F_{HH}A_{HD}F_{DD}$ results, or equivalently, the remaining missing edges in the overall ancestor graph of the univariate recursive system with independent residuals introduced at the end of Section 4.6 as a covering model to system (22). With (v) the indicator matrix $\mathcal{F}^+$ of structural zeros in $\mathcal{F}\mathcal{F}^T$ is obtained, which is at the same time the indicator matrix of structural zeros in FF^T and in $\Sigma_{(V\setminus C,V\setminus C).C}$ as implied by (25) and hence by the parent graph of A.

To illustrate the modifications directly on the edge matrix of the parent graph we add in Figure an arrow starting from node 1 and pointing to node 0, and two incoming arrows, one from node 11 pointing to node 10 and one from node 12 to node 9. We then take as conditioning set $C = \{2, 4, 5, 9\}$ and observe that its set of offspring nodes is $H = \{0, 1, 3, 6\}$ and its set of ancestor nodes outside C is $D = \{7, 8, 10, 11, 12\}$. With steps (i) and (ii) the edge matrix $\mathcal{A}_{DD}$ is modified into $\mathcal{F}_{DD}$:

$$\mathcal{A}_{DD} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathcal{F}_{DD} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where by step (i) a one is inserted in $\mathcal{A}_{DD}$ for pair (10,12). With step (iii) the edge matrix $\mathcal{A}_{HH}$ is modified into $\mathcal{F}_{HH}$:

$$\mathcal{A}_{HH} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathcal{F}_{HH} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

With step (iv) the edge matrix $\mathcal{A}_{HD}$ is modified into $\mathcal{F}_{HD}$ the structural zeros in the product $F_{HH}A_{HD}F_{DD}$:

$$\mathcal{A}_{HD} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{F}_{HD} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

By the final step (v) the upper triangular matrix edge matrix $\mathcal{F}$ of the ancestor graph in the relevant covering model of (22) is changed into $\mathcal{F}^+$, the edge matrix of the conditional covariance graph of $Y_{V \setminus C}$ given Y_C:

$$\mathcal{F}^+_{HH} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathcal{F}^+_{HD} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \mathcal{F}^+_{DD} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Especially for large graphs such an algorithmic matrix formulation may provide a much faster way of constructing the graph than using a condition involving paths. Though tracing individual paths is conceptually attractive it may be computationally somewhat tedious even in small problems.

6. Separation criteria for directed acyclic graphs

Separation criteria permit us to read directly off graphs whether the defining independence structure of the graph implies a given conditional independence statement. The problem relates for directed acyclic graphs closely to the results in the previous Sections since with $S = a \cup b$ and V partitioned into S, C, M the question to be decided is: does $a \cup b \mid C$ hold for all distributions generated over the parent graph G^V_{par} of A, that is if (2) is the defining independence structure.

Equivalence of several separation criteria: Let a, b, C, M be four nonoverlapping subsets of nodes which exhaust V and of which C or M may be empty. Then the following separation criteria for the directed acyclic graph G^V_{dag} are equivalent. Sets a and b are d-separated by C

(i) if there is no path in G^V_{dag} between a node in a and a node in b along which the following conditions hold: (1) every node with converging arrows is in C or has a descendant in C and (2) every other node is outside C (Pearl, 1988, p.117; Pearl & Verma, 1988).

(ii) if in the moral graph formed from the smallest ancestral set containing $a \cup b \cup C$ every path from a to b has a node in C (Lauritzen et al., 1990).
(This moral graph is constructed in three steps: (1) from G^V_{dag} the subgraph induced by nodes of the union of $a \cup b \cup C$ and their ancestors is obtained, (2) in it a full line is inserted for every missing (i, j)-edge having a common collision node and (3) every arrow in the resulting graph is replaced by a full line.)

(iii) if in in the parent graph G^V_{par} modified by conditioning on C there is no collisionless path from a to b wholly outside C.
(Within the subgraph of ancestors of C every missing (i, j)-edge is joined by a full line provided it has a common collision node in $C \cup$ ancestors of C.)

For a Gaussian distribution $Y_a \perp\!\!\!\perp Y_b \mid Y_C$ if the (a, b)-submatrix in the conditional covariance matrix of Y_S given Y_C is a matrix of zeros or, equivalently, if the (a, b)-submatrix in the conditional concentration matrix of Y_S given Y_C is a matrix of zeros. That is if the (a, b)-submatrices in

$$\Sigma_{SS.C} = \begin{pmatrix} \Sigma_{aa.C} & 0 \\ . & \Sigma_{bb.C} \end{pmatrix}, \quad \Sigma_{SS.C}^{-1} = \begin{pmatrix} \Sigma^{aa.M} & 0 \\ . & \Sigma^{bb.M} \end{pmatrix}$$

have structural zeros for all entries, then there is no edge between a and b in the conditional covariance graph $G_{\text{cov}}^{S.C}$ and there is no such edge in the conditional concentration graph $G_{\text{con}}^{S.C}$. By the same type of reasoning as in the previous Sections the independency, written again in node notation as $a \perp\!\!\!\perp b \mid C$, holds then not only for Gaussian dirstributions but also for general distributions generated over the same parent graph.

Therefore the three criteria are equivalent if each specifies conditions for edges between a and b to be missing in the concentration graph of Y_S given Y_C or in the covariance graph of Y_S given Y_C. Now, condition (iii) is an application of the path criterion of the previous Section for constructing edges in the covariance graph $G_{\text{cov}}^{S.C}$. The moral graph in condition (ii) is the concentration graph of nodes of S, C and the ancestors of $S \cup C$, i.e. it is $G_{\text{con}}^{V \setminus r}$, discussed in Sections 4.6 and 5.9 as an intermediate step to constructing the concentration graph $G_{\text{con}}^{S.C}$. Hence, criterion (ii) and (iii) are equivalent.

For the equivalence of the first criterion to the last we treat collisionless path and collision path separately. For collisionless paths present in the parent graph, i.e. present before conditioning on C, conditions (i) and (iii) coincide. By the last criterion a collision path between ancestors of the conditioning set is turned into a collisionless path by inserting full lines for missing (i, j)-edges provided they all have a common collision node within $L = C\cup$ ancestors of C. Since every ancestor of C has by definition a descendant in C, the relevant collision nodes are either themselves in C or they have a descendant in C. Therefore conditions (i) and (iii) coincide for collision paths as well. This completes the proof.

Further separation criteria have been studied for special types of graph by Darroch et al. (1980), Lauritzen & Wermuth (1989), Frydenberg (1990), Kauermann (1993), Spirtes (1995), Andersson et al. (1996), Koster (1996, 1999a), Studený & Bouckaert (1998), Richardson (1999) and for the more genral graphoids by Paz et al. (1999). It is conceivable that proofs can get simplified when it is tried to exploit the close connections to covariance and concentration matrices in these contexts as well.

7. Discussion

We have given a new sweep algorithm for triangular matrices and studied its relations to modifications of joint distributions generated over directed acyclic graphs. The conditions for structural zeros in matrices of linear systems turn out to be equivalent to conditions for missing edges in graphs, the main consequence being that conditions for independence statements in all Gaussian distributions generated over a given directed acyclic graph coincide with conditions for general distributions generated over the same graph.

The swt-operator is applied to a matrix of regression coefficients in an upper triangular matrix which defines together with residual variances the overall concentration matrix. However, residual variances are not used for the swt-operator nor for forming inner and outer products of relevant matrices. We therefore believe that the results apply also to semi-definite covariance structures, i.e. to degenerate joint Gaussian distributions and, more generally, to distributions without positive probability everywhere. But, we have not studied this in detail.

A number of characterizations of model subclasses can be derived as a byproduct of our results. Decomposable models are known to have a directed acyclic generating graph without a collision-oriented V-configuration (Wermuth 1980; Lauritzen and Wermuth, 1989). A further characterization is now that structural zeros in the edge matrix $\mathcal{A}$ of the parent graph coincide with structural zero in the inner product of this edge matrix, in $\mathcal{A}^T\mathcal{A}$, and hence with the edge matrix of the overall concentration graph. A conditional independence lattice model (Anderson et al. 1993) has no transition-oriented V-configuration. This is equivalent to having identical edge matrices of the parent graph and of the overall ancestor graph, i.e. to $\mathcal{A} = \mathcal{B}$. Whenever the edge matrix of the overall ancestor graph coincides with the edge matrix of the overall covariance graph, then, equivalently, the structural zeros in $\mathcal{B}$

coincide with structural zeros in its outer product $\mathcal{B}\mathcal{B}^T$ and the covariance graph is made up by nonoverlapping complete subgraphs. Finally, we have coinciding structural zeros in $\mathcal{A}$, $\mathcal{A}^T\mathcal{A}$, $\mathcal{B}$, and $\mathcal{B}\mathcal{B}^T$ if and only if all corresponding graphs consist of nonoverlapping complete subgraphs in the same components.

We believe that these types of characterization will be fruitful in studying independence equivalence of graphical Markov models in general. Many graph formulations have been given following earlier matrix results for linear systems (Wermuth, 1980; Stelzl, 1989), but a unifying result for graphs is still missing.

Appendix 1: The sweep and the resweep operator for symmetric matrices

For completeness we repeat here the definition of the original sweep and resweep operator (Beaton, 1964; Dempster, 1969)

Let M and N be a $r \times r$ symmetric matrices with elements m_{ij} and n_{ij} then with $k \neq i, k \neq j$ the operations (M swp k) and (N rswp k) are, respectively,

$$\begin{array}{llll} n_{kk} & = -1/m_{kk} & \qquad m_{kk} & = -1/n_{kk} \\ n_{ik} & = m_{ik}/m_{kk} & \qquad m_{ik} & = -n_{ik}/n_{kk} \\ n_{kj} & = m_{kj}/m_{kk} & \qquad m_{kj} & = -n_{kj}/n_{kk} \\ n_{ij} & = m_{ij} - m_{ik}m_{kj}/m_{kk}, & \qquad m_{ij} & = n_{ij} - n_{ik}n_{kj}/n_{kk}. \end{array}$$

Sweeping on all rows and columns of a set of indices a is denoted by (M swp a). The sweep operator applied to a covariance matrix Σ partitioned into three components (a, b, c) gives for instance

$$\Sigma \text{ swp } b = \begin{pmatrix} \Sigma_{aa.b} & \Pi_{a|b} & \Sigma_{ac.b} \\ . & -\Sigma_{bb}^{-1} & \Pi_{c|b}^T \\ . & . & \Sigma_{cc.b} \end{pmatrix}, \; \Sigma \text{ swp } (b, c) = \begin{pmatrix} \Sigma_{aa.bc} & \Pi_{a|b.c} & \Pi_{a|c.b} \\ . & -\Sigma_{bb.c}^{-1} & \Pi_{c|b}^T\Sigma_{cc.b}^{-1} \\ . & . & -\Sigma_{cc.b}^{-1} \end{pmatrix}.$$

By sweeping Σ on all r rows and columns $-\Sigma^{-1}$ is obtained. Resweeping $-\Sigma^{-1}$ on all r rows and columns returns Σ. The order of the sweeping operations can be interchanged without affecting the final result.

Examples of useful matrix equalities that can be directly deduced with the help of the sweep operator are $\Pi_{a|c.b} = \Sigma_{ac.b}\Sigma_{cc.b}^{-1}$, $\Pi_{a|b.c} = \Pi_{a|b} - \Pi_{a|c.b}\Pi_{c|b}$ and, after sweeping in the order c, b, by symmetry $\Pi_{a|c.b} = \Pi_{a|c} - \Pi_{a|b.c}\Pi_{b|c}$ and $\Pi_{c|b}^T\Sigma_{cc.b}^{-1} = \Sigma_{bb.c}^{-1}\Pi_{b|c}$.

Appendix 2: An equivalence of independence statements

Suppose a trivariate distribution is generated for Y_i, Y_j and Y_γ given Y_C which may be a vector variable. Suppose there is a univariate recursive generating process just as for (1) but conditionally given C, where Y_γ may be the final response, the intermediate variable or the purely explanatory corresponding to the node orderings (γ, i, j), (i, γ, j) and (i, j, γ), respectively.

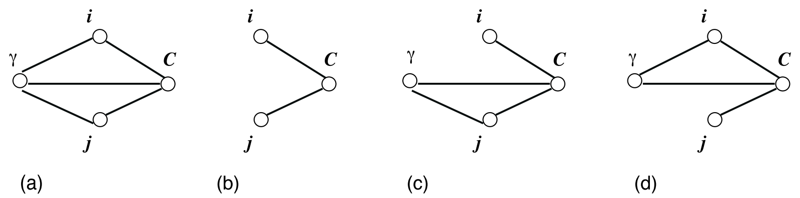
Then the following independence statements are equivalent

(i) $i \perp\!\!\!\perp j \mid (\gamma, C)$ and $i \perp\!\!\!\perp j \mid C$,

(ii) $(i, \gamma) \perp\!\!\!\perp j \mid C$ or $(j, \gamma) \perp\!\!\!\perp i \mid C$.

We do not give the argurments involving densities but just illustrate interpresta-tions in terms of concentration graphs implied by the generating process.

Suppose first that (i) holds so that in the overall concentration graph the (i, j)-edge is missing, and in the concentration graph of i, j, C alone the (i, j)-edge is also missing, i.e. when both independencies hold at the same time we may say that the independence for pair (i, j) in graph (a) is preserved after marginalising over node γ or that the independence present in graph (b) is preserved after the conditioning set is enlarged by node γ.



The left hand graph (a) alone does not imply $i \perp\!\!\!\perp j \mid C$ since there is path between i and j outside C via γ. This path via node γ vanishes if and only if either the (γ, i)-edge or the (γ, j)-edge is missing in addition, so that one of the two graphs (c) or (d) results. In graph (c) the only path from i to (γ, j) is via C, i.e. i is separated from j by C so that $(j, \gamma) \perp\!\!\!\perp i \mid C$. In graph (d) the only path from j to (γ, i) is via C so that $(i, \gamma) \perp\!\!\!\perp j \mid C$. Thus, (i) implies (ii).

Suppose next that (ii) holds, i.e. the concentration graph in the overall joint distribution is either graph (c) or graph (d). In each of these graphs every path from

i to j has a node in (γ, C) so that $i \perp\!\!\!\perp j \mid (\gamma, C)$. But, i is also separated from j by C alone, i.e. $i \perp\!\!\!\perp j \mid C$ holds as well. Thus, (ii) implies also (i).

Acknowledgement

We thank Thomas Richardson for stimulating discussions at early stages of this work and for suggesting as names: parent graph and ancestor graph.

References

Andersson, S.A. & Madigan, D. (1993). Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21**, 1318-1358.

Andersson, S.A., Madigan, D. and Perlman, M. D. (1996). An alternative Markov property for chain graphs. In Horvitz, E. and Jensen F. (eds.) *Proc. 12th Conf. Uncertainty in Artif. Intell.*, Morgan Kaufmann, San Mateo CA. 40–48.

Cochran, W.G. (1938). The omission or addition of an independent variate in multiple linear regression. *J. R. Statist. Soc.*, Suppl. **5**, 171-176.

Beaton A.E (1964). *The use of special matrix operators in statistical calculus.* Ed.D. thesis, Harvard University. Reprinted as Educational Testing Service Research Bulletin 64-51. Princeton.

Cox, D.R. & Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, **8**, 204-218; 247-277.

Cox, D.R. and Wermuth, N. (1990). An approximation to maximum-likelihood estimates in reduced models. *Biometrika*, **77**, 747-761.

Cox, D.R. & Wermuth, N. (1996). *Multivariate Dependencies – Models, Analysis and Interpretation.* London: Chapman & Hall.

Cox, D.R. & Wermuth, N. (1999). On the generation of the chordless 4-cycle. *Biometrika*, **87**, 206-212.

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980) Markov fields and log-linear models for contingency tables. *Ann. Statist.* **8**, 522–539.

Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis.* Reading: Addison-Wesley.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28**, 157-175.

Edwards, D. (1995). *Introduction to graphical modelling.* New York: Springer.

Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Statist*, **17**, 333-353.

Goldberger, A. S. (1964). *Econometric Theory.* New York: Wiley.

Jöreskog, K.G. (1981). Analysis of covariance structures. *Scan. J. Statist.* **8**, 65–92.

Kauermann, G. (1996) On a dualization of graphical Gaussian models. *Scan. J. Statist.*, to appear.

Koster, J. (1996). Markov properties of non-recursive causal models. *Ann. Statist.*, **24**,2148–2177.

Koster, J. (1999a). On the validity of the Markov interpretation of path diagrams of Gaussian structural equation systems of simultaneous equations. *Scand. J. Statist.*, **26**, 413-431.

Koster, J. (1999b). A comparison of alternative Markov properties from the point of view of applied multivariate statistics. Lecture at the workshop on 'Conditional Independence Structures and Graphical Models'. Fields Institute, Toronto.

Lauritzen, S.L. (1996). *Graphical models.* Oxford: Oxford University Press.

Lauritzen, S.L., Dawid, A.P., Larsen, B. & Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, **20**, 491-505.

Lauritzen, S.L. & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17**, 31–54.

Paz, A., Gevy, Y.R. & Studený, M. (1999). Representation of irrelevance relations by annotated graphs. To appear.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. & Verma, T. (1988). The logic of representing dependencies by directed graphs. In *Proc. Amer. Ass. Artif. Intell. Conference.* Seattle Washington.

Richardson, T. (1999). Markov properties for acyclic directed mixed graphs. *Technical Report 353*, Dept. Statistics, Univ. Washington.

Richardson, T. & Spirtes P. (2000). Ancestral graph Markov properties. *Manuscript*

Richardson, T, & Spirtes P. (2000). Ancestral graph Markov properties. *Manuscript*

Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proc. 11th conference on uncertainty and artificial intelligence* P. Besnard & S. Hanks (eds). Morgan Kaufmann: San Mateo, CA.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., & Glymour, C. (1998). Using path diagrams as a structural equation modelling tool. *Sociological Methods and Research*, **27**, 182-225.

Stelzl, I. (1989). Changing a causal hypothesis without changing the fit; some rules for generating equivalent path models. *Multivariate Behavioral Research*, **21**, 309-331.

Studený, M. & Bouckaert, R. (1998). On chain graphs models for description of conditional independence structures. *Ann. Statist.*, **26**, 1434-1495.

Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.*, **75**, 963-97.

Wermuth, N. (1992). On block-recursive linear regression equations (with discussion). *Brazilian Journal of Probability and Statistics (REBRAPE, Revista Braslileira de Probilidade e Estatistica)* , **6**, 1-56.

Wermuth, N. & Cox, D.R. (1998). On association models defined over independence graphs. *Bernoulli*, **4**, 477-495.

Wermuth, N., Cox, D.R. & Pearl J. (2000). Explanations for multivariate structures derived from univariate recursive regressions. (under revision for *Bernoulli*)

Wold, H.O.A. (1959). Ends and means in econometric model building. In (Grenander U. (ed) *Probability and Statistics.* pp.355-434. New York: Wiley.

Wold, H.O.A. & Jureen, L. (1953). Demand Analysis. New York: Wiley.

Wright, S. (1923). The theory of path coefficients: a reply to Niles' criticism. *Genetics*, **8**, 239-255.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5**, 161-215.