Partial inversion to describe linear structure

NANNY WERMUTH[1], D.R. COX[2] and M. WIEDENBECK[3]

[1]*Mathematical Statistics, Chalmers/Göteborgs Universitet, 41296 Göteborg, Sweden*
email: wermuth@math.chalmers.se

[2]*Nuffield College, Oxford, OX1, 1NF, UK*
email: david.cox@nuffield.oxford.ac.uk

[3]*Center of Survey Research, B2, 1, 68159 Mannheim, Germany*
email: wiedenbeck@zuma-mannheim.de

Abstract.

We introduce a calculus for real-valued square matrices which we call partial inversion and apply it to generating different types of statistical joint response models from a system of recursive linear equations. We also give an associated calculus for binary matrices to find what we call structural zeros and apply it to determine which elements of a set of variable pairs constrained by an independence statement before partial inversion remain so constrained after partial inversion, that is for a given set of new parameters. This permits the derivation of a wide range of different consequences of an assumed independence structure and it opens the road to compare and possibly falsify them with data for small subsets of the variables in the generating system.

AMS subject classification (2000): 00A71, 15A09, 15A23, 05C99, 62H99, 62J05.

Key words: Cholesky factorization, Directed acyclic graphs, Graphical chain model, Gauss-Doolittle method, Induced independence graphs, Successive orthogonalization, Sweep operator, Triangular systems.

1 Introduction

Methods for solving linear equations, or equivalently for matrix inversion, were known in China more than 2000 years ago [37], [22]. Into surveying, such methods have been introduced, for instance under the name of Gauss-Jordan elimination [21] in German, of Cholesky-factorization [5] in French, and of the Gauss-Doolittle method [14] in American geodesic literature. In mathematics and statistics such methods for symmetric matrices have been called successive orthogonalization [20], [33] and the sweep operator [4], [11]. Aspects of matrix decomposition and computational efficiency have been studied in numerical analysis and computer science, see e.g. [16], [36], [19]. We introduce a calculus for partial inversion of real-valued matrices, derive its properties and relate it to different types of block-triangular decompositions of positive definite matrices.

Such matrix decompositions relate closely in a statistical context to linear graphical chain models. With graphical chain models [9], [15], [25], [39] one can formulate relations among many random variables of arbitrary distributional form, such that simplifying structure results from conditional independencies and is captured in graphs in which nodes represent random variables and missing edges indicate sets of parameters constrained to take value zero, capturing independence. In such systems, there is typically a direction of dependence between some but not all pairs of nodes.

These multivariate statistical models combine three essential concepts which have been developed independently at the beginning of the 20th century. The geneticist Wright [43] used directed graphs to formulate hypotheses how linear relations in his data could have been generated. The probabilist Markov [28] introduced the notion of conditional independence to represent seemingly complex structures by a sequence of univariate dependencies, and the physicist Gibbs [18] characterized the higher density of a substance by an undirected graph, in which nodes have a larger number of nearest neighbors. Many properties and estimation algorithms for different subclasses of graphical models have been established in the statistical literature in the last 30 years, but these have so far not been connected to properties of matrix operators.

Key questions in statistical modeling are: how is the strength and direction of dependence between two random variables changed when their set of conditioning variables is modified, and in which situations are both properties preserved. These questions concern the parameters of a model which are free to vary, often within some range of non-vanishing dependence. Answers are essential for comparing results of different empirical studies on the same set of core variables. In two studies of even the same set of variables, different sequences for the variables may be used for analysis or only a partial ordering be given, since some variables are to be considered as joint responses. Or, it may be that some variables are omitted, i.e. marginalized over, or a sub-population is studied for which some levels of other variables are held fixed, i.e. are conditioned on.

Closely related are inquiries into change and preservation of independence constraints specified by a given graph. For this, we introduce a calculus for finding structural zeros after partial inversion, derive its properties and apply it to graphical chain models. This calculus operates on binary matrices. Many preserved independencies typically simplify statistical analysis even when the interpretation of the constraints and of the unconstrained parameters is changed.

Necessary for the new matrix results are a minor modification of the sweep operator, so that it becomes applicable to real valued, square matrices instead of only to symmetric matrices, and a minor modification of adjacency matrices, the binary matrix representations traditional in the graph theoretic literature, so that matrix products of the new binary matrices, called edge matrices, become analogous to the real-valued matrix products in partial inversion.

The plan of the paper is to introduce partial inversion in Section 2. In Section

3, partial inversion is applied to symmetric matrices and related to the statistical concept of linear least squares regression coefficients, to conditional covariance matrices and to inverse marginal covariance matrices. Different properties of these types of parameters motivate the use of independence graphs with several types of edge. In Section 4, the discussion is extended to parameter and edge matrices of linear graphical chain models induced after partial inversion. In a short final section we point to some open problems.

2 Partial inversion and its properties

2.1 Definition and basic properties

Let $N = \{1, \dots, d_N\}$ be the index set of rows and corresponding columns of a square matrix M, whose principal submatrices are all invertible, i.e. for which the inverse of $M_{aa} = [M]_{a,a}$ exists for every nonempty subset a of N, and is denoted by M_{aa}^{-1}. Let further N be split into two arbitrary components a and b, so that, if necessary after permuting rows and columns, we get $N = (a, b)$. For two real valued vectors x and y split accordingly, we are to introduce below in equation (2.2) an operation on M, to be called partial inversion and denoted by inv_a, such that for

$$M \begin{pmatrix} x_a \\ x_b \end{pmatrix} = \begin{pmatrix} y_a \\ y_b \end{pmatrix}$$

the linear relation after partial inversion is

$$\mathrm{inv}_a M \begin{pmatrix} y_a \\ x_b \end{pmatrix} = \begin{pmatrix} x_a \\ y_b \end{pmatrix}. \tag{2.1}$$

For this we write M and its inverse M^{-1} in partitioned form as

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

Partial inversion of M on subset a of N and a convenient notation are then defined by

$$\mathrm{inv}_a M = \begin{pmatrix} M_{aa}^{-1} & -M_{aa}^{-1} M_{ab} \\ M_{ba} M_{aa}^{-1} & M_{bb} - M_{ba} M_{aa}^{-1} M_{ab} \end{pmatrix} = \begin{pmatrix} M_{aa}^{-1} & -M_{a \leftarrow b} \\ M_{b \rightarrow a} & M_{bb.a} \end{pmatrix}. \tag{2.2}$$

The notation $M_{a \leftarrow b}$ reminds us that the matrix M_{ab} is multiplied from the left by M_{aa}^{-1} and $M_{bb.a}$ is the notation for what is often called a Schur matrix. That the operator defined in equation (2.2) is of the desired form in equation (2.1) is verified by writing the component for a from $Mx = y$ as

$$y_a = M_{aa} x_a + M_{ab} x_b$$

and substituting it on the left-hand side of equation (2.1).

One use of partial inversion is to decompose matrix inversion into steps of the same kind, which lead to the inverse of M if applied in sequence to each element of N. In the following we study the properties of this operator, derive several recursion relations from it and show some of its applications to linear statistical models, to matrix representations of graphs. Some of the results hold under weaker assumptions, but our main applications concern positive definite matrices and unit triangular matrices, i.e. triangular matrices having ones along the diagonal, for both of which all principal submatrices are invertible.

LEMMA 2.1. *Basic properties of partial inversion.*
(i) Partial inversion on component a is undone by reapplying it to a:

$$\text{inv}_a \text{inv}_a M = M;$$

(ii) the matrix M partially inverted on a is the inverse of M after partial inversion on the remaining components b:

$$\text{inv}_a M = (\text{inv}\,_b M)^{-1};$$

(iii) partial inversion on component a of M followed by partial inversion on the remaining components b gives the inverse of M:

$$\text{inv}\,_b \text{inv}_a M = M^{-1};$$

(iv) the order of partial inversion on components a and b can be interchanged:

$$\text{inv}\,_b \text{inv}_a M = \text{inv}_a \text{inv}\,_b M;$$

(v) the matrix M partially inverted on a coincides with its inverse M^{-1} partially inverted on the remaining components b:

$$\text{inv}_a M = \text{inv}\,_b M^{-1};$$

(vi) inversion and partial inversion can be interchanged:

$$(\text{inv}\,_b M)^{-1} = \text{inv}\,_b M^{-1} \ .$$

PROOF. Properties (i), (iii) and (iv) are direct from equation (2.1), property (v) results with property (i) applied to partial inversion on b in (iii), and (vi) is direct from (ii) and (v).
For property (ii) equality results from equation (2.1) and

$$\text{inv}_b M \begin{pmatrix} x_a \\ y_b \end{pmatrix} = \begin{pmatrix} y_a \\ x_b \end{pmatrix}, \quad (\text{inv}_b M)^{-1} \begin{pmatrix} y_a \\ x_b \end{pmatrix} = \begin{pmatrix} x_a \\ y_b \end{pmatrix}$$

and the proof is complete.

It is, however, instructive to prove the basic properties of the operator also directly by matrix calculations. Then, case (i) results by applying partial inversion to component a of the matrix $\mathrm{inv}_a M$ in equation (2.2). Direct computation gives property (ii), with

$$\begin{pmatrix} M_{aa}^{-1} & -M_{a\leftharpoondown b} \\ M_{b\rightharpoondown a} & M_{bb.a} \end{pmatrix} \begin{pmatrix} M_{aa.b} & M_{a\rightharpoondown b} \\ -M_{b\leftharpoondown a} & M_{bb}^{-1} \end{pmatrix} = \begin{pmatrix} I_{aa} & 0_{ab} \\ 0_{ba} & I_{bb} \end{pmatrix}.$$

Partial inversion on component b applied to $\mathrm{inv}_a M$ in equation (2.2) can be written as

$$\text{(2.3)} \quad \mathrm{inv}_b \mathrm{inv}_a M = \mathrm{inv}_N M = \begin{pmatrix} M_{aa}^{-1} + M_{a\leftharpoondown b} M_{bb.a}^{-1} M_{b\rightharpoondown a} & -M_{a\leftharpoondown b} M_{bb.a}^{-1} \\ -M_{bb.a}^{-1} M_{b\rightharpoondown a} & M_{bb.a}^{-1} \end{pmatrix}$$

and

$$(M_{aa} \quad M_{ab}) \mathrm{inv}_N M = (I_{aa} \quad 0_{ab}), \quad (-M_{bb.a}^{-1} M_{b\rightharpoondown a} \quad M_{bb.a}^{-1}) M = (0_{ba} \quad I_{bb})$$

proving that M^{-1} has the form of the matrix in equation (2.3), so that property (iii) holds. The exchangeability in case (iv) follows from case (iii) and a as well as b being arbitrary nonempty subsets of N. □

2.2 Directly related results

Partial inversion in equation (2.2) generalizes the sweep operator for symmetric matrices of Beaton [4] for which Dempster [11], [12] has shown that it has properties (iii) to (v). The remaining properties of Lemma 2.1 do not hold since sweeping is undone by resweeping, which is similar but not identical to sweeping. The sweep operator differs from equation (2.2) by the minus sign in the upper part and gives $-M^{-1}$ after sweeping on N while partial inversion on N gives M^{-1}. If N is partitioned into more than two components the definition of both these operators still applies component by component.

Property (v), written here explicitly for partial inversion of M on b, is a standard equality for partitioned inverses:

$$\text{(2.4)} \qquad \begin{pmatrix} M_{aa.b} & M_{a\rightharpoondown b} \\ -M_{b\leftharpoondown a} & M_{bb}^{-1} \end{pmatrix} = \begin{pmatrix} (M^{aa})^{-1} & -(M^{aa})^{-1} M^{ab} \\ -M^{ba}(M^{aa})^{-1} & M^{bb.a} \end{pmatrix}.$$

Compact explicit forms for the partitioned inverse result from equation (2.3) and the basic properties (iv) and (v) of partial inversion, such as

$$\text{(2.5)} \qquad M^{aa} = M_{aa.b}^{-1}, \quad -M^{ab} = M_{aa.b}^{-1} M_{a\rightharpoondown b} = M_{a\leftharpoondown b} M_{bb.a}^{-1}.$$

Equation (2.5) permits us to introduce some further notation for components of partially inverted matrices. For instance, for $N = \{G, J\}$ with $G = \{a, b\}$ and $J = \{c, d\}$, component (G, d) in $M_{G\leftharpoondown d} = M_{GG}^{-1} M_{Gd}$ is

$$
\begin{pmatrix} M_{aa.b}^{-1} & -M_{aa.b}^{-1}M_{a\rightharpoondown b} \\ -M_{bb.a}^{-1}M_{b\rightharpoondown a} & M_{bb.a}^{-1} \end{pmatrix} \begin{pmatrix} M_{ad} \\ M_{bd} \end{pmatrix} = \begin{pmatrix} M_{aa.b}^{-1}M_{ad.b} \\ M_{bb.a}^{-1}M_{bd.a} \end{pmatrix} = \begin{pmatrix} M_{a\leftharpoondown d.b} \\ M_{b\leftharpoondown d.a} \end{pmatrix} \tag{2.6}
$$

and component (d, G) is

$$
M_{d\rightharpoondown G} = \left(\; M_{da.b}M_{aa.b}^{-1} \quad M_{db.a}M_{bb.a}^{-1} \; \right) = \left(\; M_{d.b\rightharpoondown a} \quad M_{d.a\rightharpoondown b} \; \right), \tag{2.7}
$$

where, e.g., $M_{ad.b} = M_{ad} - M_{ab}M_{bb}^{-1}M_{bd}$. The notation $M_{b\leftharpoondown d.a}$ reminds us that partial inversion has first been carried out on a, then by additional partial inversion on b, the component (b, d) is the matrix $M_{bd.a}$ multiplied to the left by $M_{bb.a}^{-1}$. Similarly, $M_{d.a\rightharpoondown b}$ is the matrix $M_{db.a}$ multiplied to the right by $M_{bb.a}^{-1}$.

For $N = \{a, K\}$ with $K = \{b, c, d\}$, the change of partial inversion on a to partial inversion on $\{a, b\} = G$ can now be studied in Theorem 2.2 and Corollary 2.3 below by using the following types of compact expressions for the resulting matrices

$$
\mathrm{inv}_a M = \left(\begin{array}{c|ccc} M_{aa}^{-1} & -M_{a\leftharpoondown b} & -M_{a\leftharpoondown c} & -M_{a\leftharpoondown d} \\ \hdashline M_{b\rightharpoondown a} & M_{bb.a} & M_{bc.a} & M_{bd.a} \\ M_{c\rightharpoondown a} & M_{cb.a} & M_{cc.a} & M_{cd.a} \\ M_{d\rightharpoondown a} & M_{db.a} & M_{dc.a} & M_{dd.a} \end{array} \right), \tag{2.8}
$$

and, by using equation (2.5), we get

$$
\mathrm{inv}_G M = \left(\begin{array}{cc|cc} M_{aa.b}^{-1} & -M_{a\leftharpoondown b}M_{bb.a}^{-1} & -M_{a\leftharpoondown c.b} & -M_{a\leftharpoondown d.b} \\ -M_{bb.a}^{-1}M_{b\rightharpoondown a} & M_{bb.a}^{-1} & -M_{b\leftharpoondown c.a} & -M_{b\leftharpoondown d.a} \\ \hdashline M_{c.b\rightharpoondown a} & M_{c.a\rightharpoondown b} & M_{cc.ab} & M_{cd.ab} \\ M_{d.b\rightharpoondown a} & M_{d.a\rightharpoondown b} & M_{dc.ab} & M_{dd.ab} \end{array} \right). \tag{2.9}
$$

2.3 Main derived properties

Now the main properties of partial inversion can be summarized.

THEOREM 2.2. *Commutativity, exchangeability and symmetric difference for partial inversion.* Let arbitrary components a, b, c partition N, $G = \{a, b\}$, and the matrix M be accordingly partitioned, then

$$
\begin{aligned}
&(i) \quad \mathrm{inv}_a\, \mathrm{inv}_b M = \mathrm{inv}_b\, \mathrm{inv}_a M = \mathrm{inv}_{ab} M; \\
&(ii) \quad [\mathrm{inv}_a M]_{G,G} = \mathrm{inv}_a M_{GG}; \\
&(iii) \; \mathrm{inv}_{ab}\, \mathrm{inv}_{bc} M = \mathrm{inv}_{ac} M\,.
\end{aligned}
$$

PROOF. The commutativity in case (i) results with the change from equation (2.8) to equation (2.9) and properties (iii) and (iv) in Lemma 2.1. The exchangeability in case (ii) of a submatrix chosen after partial inversion and partial inversion carried out on a submatrix is a consequence of property (i) in this Theorem and of the definition in equation (2.2) of the operator. The symmetric difference property (iii) results from the cases (i) both in this Theorem and in Lemma 2.1. □

ILLUSTRATION 2.1. For a square matrix M of dimension $d_N = 3$ and elements m_{ij} partial inversion on $a = \{1, 2\}$, carried out in two steps by starting with row and column 1, gives directly from equation (2.2)

$$\text{inv}_1 M = \left(\begin{array}{c|cc} 1/m_{11} & -m_{12}/m_{11} & -m_{13}/m_{11} \\ \hdashline m_{21}/m_{11} & m_{22.1} & m_{23.1} \\ m_{31}/m_{11} & m_{32.1} & m_{33.1} \end{array}\right).$$

Partially inverting this matrix on row and column 2, using equation (2.5) and exploiting case (i) of Theorem 2.2 gives the compact expression

$$\text{inv}_{12} M = \left(\begin{array}{cc|c} 1/m_{11.2} & -m_{12}/(m_{11}m_{22.1}) & -m_{13.2}/m_{11.2} \\ -m_{21}/(m_{11}m_{22.1}) & 1/m_{22.1} & -m_{23.1}/m_{22.1} \\ \hdashline m_{31.2}/m_{11.2} & m_{32.1}/m_{22.1} & m_{33.12} \end{array}\right).$$

2.4 Recursion relations

General recursion relations result from explicit expressions of the matrices obtained after successive steps of partial inversion.

COROLLARY 2.3. *Recursion relations obtained by partial inversion.* Let the subsets a, b, c, d partition N and the matrix M be accordingly partitioned. Let further $G = \{a, b\}$, $H = \{a, b, c\}$, $J = \{c, d\}$ and $K = \{b, c, d\}$. Then, when M is partially inverted in sequence on a, b, c, there results

(i) for elements corresponding to (c, d) in $M_{KK.a}$ and in $M_{JJ.G}$

$$M_{cd.a} = M_{cd} - M_{ca}M_{aa}^{-1}M_{ad}, \quad M_{cd.ab} = M_{cd.a} - M_{cb.a}M_{bb.a}^{-1}M_{bd.a}\,;$$

(ii) for elements corresponding to (a, d) in $M_{G\leftharpoondown J}$ and in $M_{H\leftharpoondown d}$

$$M_{a\leftharpoondown d.b} = M_{a\leftharpoondown d} - M_{a\leftharpoondown b}M_{b\leftharpoondown d.a}, \quad M_{a\leftharpoondown d.bc} = M_{a\leftharpoondown d.b} - M_{a\leftharpoondown c.b}M_{c\leftharpoondown d.ab}\,;$$

(iii) for elements corresponding to (d, a) in $M_{J\rightharpoondown G}$ and in $M_{d\rightharpoondown H}$

$$M_{d.b\rightharpoondown a} = M_{d\rightharpoondown a} - M_{d.a\rightharpoondown b}M_{b\rightharpoondown a}, \quad M_{d.bc\rightharpoondown a} = M_{d.b\rightharpoondown a} - M_{d.ab\rightharpoondown c}M_{c.b\rightharpoondown a};$$

(iv) for elements corresponding to (a, a) in M_{GG}^{-1} and in M_{HH}^{-1}

$$M_{aa.b}^{-1} = M_{aa}^{-1} - M_{a\leftharpoondown b}M_{bb.a}^{-1}M_{b\rightharpoondown a}, \quad M_{aa.bc}^{-1} = M_{aa.b}^{-1} - M_{a\leftharpoondown c.b}M_{cc.ab}^{-1}M_{c.b\rightharpoondown a}\ .$$

Proof. The relations result by interpreting the modifications due to repeated partial inversion. The forms of $M_{cd.ab}$, $M_{a\leftharpoondown d.b}$, $M_{d.b\rightharpoondown a}$, and $M^{-1}_{aa.b}$ are components when changing from $\mathrm{inv}_a M$ in equation (2.8) to $\mathrm{inv}_b \mathrm{inv}_a M$ in equation (2.9). Similarly, $M_{a\leftharpoondown d.bc}$, $M_{d.bc\rightharpoondown a}$, and $M^{-1}_{aa.bc}$ are components when changing from $\mathrm{inv}_G M$ in equation (2.9) to $\mathrm{inv}_c \mathrm{inv}_G M$. □

2.5 Directly related matrix decompositions

Direct computations show also that rows and columns of the matrices in a block-triangular decomposition $M = \mathsf{L}K\urcorner$, where K is a block-diagonal matrix, L is a unit lower block-triangular matrix and $\urcorner$ is a unit upper block-triangular matrix, can be specified in terms of partial inversion. For the following Lemma and throughout the paper let d arbitrary subsets of N be given. When these are ordered as $(1, \ldots, g, \ldots, d)$ then we call the result an ordered partitioning of N.

Lemma 2.4. *The relation of block-triangular decompositions to partial inversion.* Let an ordered partitioning $(1, \ldots, g, \ldots, d)$ of N be given, and $M = \mathsf{L}K\urcorner$ be a corresponding block-triangular decomposition. Let r denote all indices to the right and l all indices to the left of g, both excluding g. Furthermore, let L_{Ng} be columns of L and $\urcorner_{gN}$ be the rows of $\urcorner$ corresponding to g. Then

$$\mathsf{L}_{Ng} = \begin{pmatrix} 0_{lg} \\ I_{gg} \\ M_{r.l\rightharpoondown g} \end{pmatrix}, \quad K_{gg} = M_{gg.l}, \quad \urcorner_{gN} = (0_{gl} \ I_{gg} \ M_{g\leftharpoondown r.l}). \tag{2.10}$$

Thereby we use the convention that the submatrix of indices (l, g) is absent when $g = 1$ and the submatrix of (g, r) is absent when $g = d$.

Illustration 2.2. For instance, for $d = 3$ and the three blocks denoted by a, b, c, the decomposition is

$$M = \begin{pmatrix} I_{aa} & 0 & 0 \\ M_{b\rightharpoondown a} & I_{bb} & 0 \\ M_{c\rightharpoondown a} & M_{c.a\rightharpoondown b} & I_{cc} \end{pmatrix} \begin{pmatrix} M_{aa} & 0 & 0 \\ 0 & M_{bb.a} & 0 \\ 0 & 0 & M_{cc.ab} \end{pmatrix} \begin{pmatrix} I_{aa} & M_{a\leftharpoondown b} & M_{a\leftharpoondown c} \\ 0 & I_{bb} & M_{b\leftharpoondown c.a} \\ 0 & 0 & I_{cc} \end{pmatrix}.$$

The form of M with the partitioning refined so that each block contains a single element leads to the following result which is closely related to those given recently under slightly weaker assumptions [19].

Lemma 2.5. *Decomposition of M into a symmetric and a unit triangular matrix.* A square matrix M, whose principal submatrices are all invertible, can be decomposed into the invertible symmetric matrix $S = \mathsf{L}K\mathsf{L}^{\mathrm{T}}$ and the unit upper-triangular matrix $\urcorner = \mathsf{L}^{-\mathrm{T}}\urcorner$ chosen so that $M = S\urcorner$. In addition, there is the decomposition $S^* = \urcorner^{\mathrm{T}} K \urcorner$ and the unit lower-triangular matrix $\mathsf{L} = \mathsf{L}\urcorner^{-\mathrm{T}}$ chosen so that $M = \mathsf{L}S^*$, where L, $\urcorner$ and K are given by equation (2.10) for $d = d_N$ and e.g. $\mathsf{L}^{-\mathrm{T}}$ denotes the transpose of the inverse of L.

PROOF. For $d = d_N$, the matrix L in the decomposition given in equation (2.10) is unit lower triangular, K is diagonal, and $\urcorner$ is unit upper-triangular. The inverse $\mathsf{L}^{-\mathrm{T}}$ of L^{T} is unit upper-triangular and the product of two unit upper-triangular matrices is of the same form, so that the symmetric matrix $S = \mathsf{L}K\mathsf{L}^{\mathrm{T}}$ has the determinant of M and there is a similar argument for $\urcorner$. $\square$

The decomposition of equation (2.10) applied to a symmetric matrix M has $\urcorner = \mathsf{L}^{\mathrm{T}}$, so that $M = \mathsf{L}K\urcorner$ in Lemma 2.5 reduces with $S = S^*$ to the usual triangular decomposition of a positive definite matrix. There is also the interpretation of L^{-1} as block-triangularizing M from the left and of $\urcorner^{-1}$ as block-triangularizing M from the right with

$$(2.11) \qquad {}'M = \mathsf{L}^{-1}M = K\urcorner, \quad M' = M\urcorner^{-1} = \mathsf{L}K,$$

so that $'M$ is upper block-triangular and M' is lower block-triangular.

Now the elements of the block-triangular and triangular decompositions of an invertible symmetric matrix permit an interpretation in terms of partial inversion, which is to be given in Theorem 4.2 in the following section.

3 Some direct applications to linear models

3.1 Elements of partially inverted covariance matrices

Let Σ be the invertible covariance matrix of a mean-centered column vector random variable Y, and let Σ^{-1} be the concentration matrix of Y. Let further a split of Y be defined by two arbitrary vector components, Y_a and Y_b. Then, Lemma 2.1 (v) equates two nonsymmetric matrices

$$(3.1) \qquad \mathrm{inv}_a\Sigma^{-1} = \begin{pmatrix} (\Sigma^{aa})^{-1} & -(\Sigma^{aa})^{-1}\Sigma^{ab} \\ \sim & \Sigma^{bb.a} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa|b} & \Sigma_{ab}\Sigma_{bb}^{-1} \\ \sim & \Sigma_{bb}^{-1} \end{pmatrix} = \mathrm{inv}_b\Sigma.$$

Here and throughout, the $\sim$ notation indicates entries in a matrix which is symmetric up to the sign, i.e. minus elements given in the upper off-diagonal part of the matrix.

The off-diagonal matrices specify two different, but equivalent, ways of computing $\Pi_{a|b}$, the matrix of regression coefficients of Y_b in linear least squares regression of Y_a on Y_b [10]. This coefficient matrix is defined by the linear equation $Y_a = \Pi_{a|b}Y_b + \varepsilon_a$ in which $\mathrm{cov}(\varepsilon_a, Y_b^{\mathrm{T}}) = 0$, i.e. with

$$(3.2) \qquad \Sigma_{ab} = E(Y_a\, Y_b^{\mathrm{T}}) = \Pi_{a|b}E(Y_b\, Y_b^{\mathrm{T}}) + E(\varepsilon_a Y_b^{\mathrm{T}}) = \Pi_{a|b}\Sigma_{bb}.$$

The interpretation of $\Sigma^{bb.a} = \Sigma_{bb}^{-1}$ in equation (3.1) as the concentration matrix of Y_b and of $(\Sigma^{aa})^{-1} = \Sigma_{aa|b}$ as the covariance matrix of $Y_{a|b} = Y_a - \Pi_{a|b}Y_b$ had been derived by Dempster [11] in terms of the sweep operator for symmetric matrices. To distinguish Schur matrices resulting in this context by marginalizing and by conditioning, we use the notation $\Sigma^{bb.a}$ and $\Sigma_{aa|b}$.

With subsets a, b, c, d partitioning N, $G = \{a, b\}$ and $J = \{c, d\}$ we denote the different components of least squares regression coefficient matrices by

$$\Pi_{G|J} = (\Pi_{G|c.d}\ \Pi_{G|d.c}), \quad \Pi_{G|J} = \begin{pmatrix} \Pi_{a|J} \\ \Pi_{b|J} \end{pmatrix},$$

so that e.g. $\Pi_{b|c.d}$ contains the coefficients of Y_c in a least squares regression of Y_b on both Y_c and Y_d. Matrix forms of recursion relations for least squares regression coefficients [6], for covariances [1], and for concentrations [11], are then recognized to be consequences of the recursion properties of partial inversion and are

$$\begin{aligned} \Pi_{b|d.c} &= \Pi_{b|d} - \Pi_{b|c.d}\Pi_{c|d}\,; \\ \Sigma_{bd|c} &= \Sigma_{bd} - \Sigma_{bc}\Sigma_{cc}^{-1}\Sigma_{cd}\,; \\ \Sigma^{bd.a} &= \Sigma^{bd} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ad}. \end{aligned} \tag{3.3}$$

Equations (3.3) provide insight into when regression coefficients, covariances and concentrations remain unchanged after marginalizing or after conditioning. Elements in positions (i, j) of $\mathrm{inv}_G\Sigma^{-1}$ are proportional to different types of partial correlation, compare e.g. [1], Sections 2.3, 2.5, [9], Section 3.4, denoted here by

$$\begin{aligned} &(i) \quad \rho_{ij|J\setminus j} \text{ for } \Pi_{G|J} \text{ with } i \in G,\ j \in J\,; \\ &(ii) \quad \rho_{ij|J} \text{ for } \Sigma_{GG|J} \text{ with } i \in G,\ j \in G\,; \\ &(iii)\ \rho_{ij|J\setminus\{ij\}} \text{ for } \Sigma^{JJ.G} \text{ with } i \in J,\ j \in J\,. \end{aligned} \tag{3.4}$$

Equations (3.4) give the interpretation of different types of edge to represent an independence structure in $\mathrm{inv}_G\Sigma^{-1}$, arrows for $\Pi_{G|J}$ and two different types of undirected graph for $\Sigma_{GG|J}$, called a conditional covariance graph, and in $\Sigma^{JJ.G}$, called a marginal concentration graph.

Equations (3.4) and (3.3) jointly capture also how probabilistic independence statements combine for Gaussian distributions that are nondegenerate, i.e. for which Σ is positive definite. With Y_b independent of Y_d given Y_c denoted by $b \perp\!\!\!\perp d|c$,

$$\begin{aligned} &(i) \quad b \perp\!\!\!\perp d \text{ and } (b \perp\!\!\!\perp c|d \text{ or } c \perp\!\!\!\perp d) \text{ imply } b \perp\!\!\!\perp d|c\,; \\ &(ii) \quad b \perp\!\!\!\perp d \text{ and } (b \perp\!\!\!\perp c \text{ or } c \perp\!\!\!\perp d) \text{ imply } b \perp\!\!\!\perp d|c\,; \\ &(iii)\ b \perp\!\!\!\perp d|ac \text{ and } (b \perp\!\!\!\perp a|cd \text{ or } d \perp\!\!\!\perp a|bc) \text{ imply } b \perp\!\!\!\perp d|c\,. \end{aligned} \tag{3.5}$$

3.2 Relations to triangular decompositions of Σ and of Σ^{-1}

Partial inversion applied repeatedly to symmetric matrices leads in particular to the following interpretations of the resulting matrix components.

THEOREM 3.1. *Interpretation of block-triangular decompositions of invertible Σ and Σ^{-1}.* Let an ordered partitioning $(1, \ldots, g, \ldots, d)$ of N be given. Let further r denote all indices to the right and l all indices to the left of g excluding g in this new ordering, and the unit block-triangular decompositions be $\Sigma^{-1} = \urcorner^{\mathrm{T}} H \urcorner$, compare Illustration (2.2), and $\Sigma = \mathsf{L}^{\mathrm{T}} K \mathsf{L}$. Then

$$(i)\;\; H = K^{-1} \text{ and } \mathsf{L}^{\mathrm{T}} = \urcorner^{-1};$$
$$(ii)\; H_{gg} = \Sigma^{gg.l}, \quad \urcorner_{gN} = (0_{gl}\; I_{gg}\; -\Pi_{g|r});$$
$$(iii) K_{gg} = \Sigma_{gg|r}, \; \mathsf{L}_{gN} = (\Pi^{\mathrm{T}}_{l|g.r}\; I_{gg}\; 0_{gr}),$$

where $\Sigma^{gg.l}$ and $\Sigma_{gg|r}$ are the concentration matrix and the covariance matrix of $Y_{g|r}$.

ILLUSTRATION 3.1. For instance, for $d = 4$, the lower block-triangular decomposition $(H, \urcorner)$ of Σ^{-1} is

$$(3.6)\;\; H = \begin{pmatrix} \Sigma^{aa} & 0 & 0 & 0 \\ . & \Sigma^{bb.a} & 0 & 0 \\ . & . & \Sigma^{cc.ab} & 0 \\ . & . & . & \Sigma^{dd.abc} \end{pmatrix}, \quad \urcorner = \begin{pmatrix} I_{aa} & -\Pi_{a|b.cd} & -\Pi_{a|c.bd} & -\Pi_{a|d.bc} \\ 0 & I_{bb} & -\Pi_{b|c.d} & -\Pi_{b|d.c} \\ 0 & 0 & I_{cc} & -\Pi_{c|d} \\ 0 & 0 & 0 & I_{dd} \end{pmatrix},$$

where, here, and throughout, the . notation indicates entries in a symmetric matrix, i.e. elements given in the upper off-diagonal part of the matrix. For the block-triangular decomposition $(K, \mathsf{L}^{\mathrm{T}})$ of Σ we have

$$(3.7)\;\; K = \begin{pmatrix} \Sigma_{aa|bcd} & 0 & 0 & 0 \\ . & \Sigma_{bb|cd} & 0 & 0 \\ . & . & \Sigma_{cc|d} & 0 \\ . & . & . & \Sigma_{dd} \end{pmatrix}, \quad \mathsf{L} = \begin{pmatrix} I_{aa} & 0 & 0 & 0 \\ \Pi^{\mathrm{T}}_{a|b.cd} & I_{bb} & 0 & 0 \\ \Pi^{\mathrm{T}}_{a|c.d} & \Pi^{\mathrm{T}}_{b|c.d} & I_{cc} & 0 \\ \Pi^{\mathrm{T}}_{a|d} & \Pi^{\mathrm{T}}_{b|d} & \Pi^{\mathrm{T}}_{c|d} & I_{dd} \end{pmatrix}.$$

PROOF. Case (i) of Theorem (3.1) is direct by matrix inversion. Case (ii) is proven by Lemma 2.4. Furthermore, since the concentration matrix of Y_b has with $\Sigma^{bb.a}$ the same form as Σ^{-1}, the same type of argument applies to it and, similarly, to $\Sigma^{gg.l}$. In this way, the block-triangular decomposition of equation (2.10) is built up for Σ^{-1}. To prove case (iii), we note that the form of the block-triangular decomposition of Σ results by matrix inversion from case (ii) and by the recursion relations for regression coefficients in equation (3.3). □

3.3 Relations to triangularized forms of Σ and of Σ^{-1}

With a given ordered partitioning $N = (1, \ldots, g, \ldots, d)$ block g is associated with the vector variable Y_g. For such a sequence of vector variables we denote the left-triangularized form of the concentration matrix by $'\Sigma^{-1}$ and the right-triangularized form of the covariance matrix by Σ', and next derive their components before and after partial inversion.

COROLLARY 3.2. *Interpretation of partial inversion for* ${}'\Sigma^{-1}$ *and* Σ'. Let the matrices ${}'\Sigma^{-1} = H\urcorner$ and $\Sigma' = K\mathsf{L}$ be defined from the block-triangular decompositions of Theorem 4.2 for an invertible covariance matrix Σ. Then

$$\begin{aligned}
&(i)\ \ [{}'\Sigma^{-1}]_{g,N} = (0_{gl}\ \ \Sigma^{gg.l}\ \ \Sigma^{gr.l});\\
&(ii)\ \Sigma'_{gN} = (\Sigma_{gl|r}\ \ \Sigma_{gg|r}\ \ 0_{gr});\\
&(iii)[\mathrm{inv}_{lg}{}'\Sigma^{-1}]_{g,N} = (0_{gl}\ \ \Sigma_{gg|r}\,\Pi_{g|r}),
\end{aligned}$$

where $\Sigma^{gr.l}$ gives the concentrations of Y_g and Y_r, while $\Sigma_{gl|r}$ gives the covariances of $Y_{g|r}$ and $Y_{l|r}$; all other submatrices are as defined for Theorem 4.2.

ILLUSTRATION 3.2. For $d = 4$, the left block-triangularized matrix ${}'\Sigma^{-1}$ in (i) and the right-triangularized matrix Σ' in (ii) are

$$ {}'\Sigma^{-1} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} & \Sigma^{ad}\\ 0 & \Sigma^{bb.a} & \Sigma^{bc.a} & \Sigma^{bd.a}\\ 0 & 0 & \Sigma^{cc.ab} & \Sigma^{cd.ab}\\ 0 & 0 & 0 & \Sigma^{dd.abc} \end{pmatrix},\ \Sigma' = \begin{pmatrix} \Sigma_{aa|bcd} & 0 & 0 & 0\\ \Sigma_{ba|cd} & \Sigma_{bb|cd} & 0 & 0\\ \Sigma_{ca|d} & \Sigma_{cb|d} & \Sigma_{cc|d} & 0\\ \Sigma_{da} & \Sigma_{db} & \Sigma_{dc} & \Sigma_{dd} \end{pmatrix}. $$

The matrices of case (iii) are, for this example of four blocks, the matrices K and $\urcorner$ in equations (3.6) and (3.7).

PROOF. From the block-triangular decomposition $\Sigma^{-1} = \urcorner^{\mathrm{T}} H \urcorner$, where $\urcorner$ is unit upper block-triangular, the matrix ${}'\Sigma^{-1} = H\urcorner$ is upper block- triangular. The interpretation in (i) follows from the product $H\urcorner$ by using the definition of least-squares regression coefficient matrices in terms of concentrations. A similar argument applies to the block-triangular decomposition $\Sigma = \mathsf{L}^{\mathrm{T}} K \mathsf{L}$, where $\mathsf{L} = \urcorner^{-\mathrm{T}}$ is unit lower block-triangular by using the definition of least-squares regression coefficient matrices in terms of covariances for the product $K\mathsf{L}$.

The upper block-triangularity of ${}'\Sigma^{-1}$, the definition of partial inversion in equation (2.2) and the exchangeability property in Theorem 4.2 case (ii) imply for $\bar{l} = N \setminus l$ that

$$[\mathrm{inv}_{1,\dots,g}{}'\Sigma^{-1}]_{\bar{l},\bar{l}} = \mathrm{inv}_g[{}'\Sigma^{-1}]_{\bar{l},\bar{l}} = \begin{pmatrix} \Sigma_{gg|r} & \Pi_{g|r}\\ 0_{rg} & {}'\Sigma^{rr.lg} \end{pmatrix},$$

and hence the form given in (iii). □

4 Relations to linear graphical chain models

The partial inversion results in the previous section relate directly to linear stepwise data generating processes and to different types of statistical joint response models which can be generated by them.

4.1 Relations to linear triangular systems

With a partitioning of N refined to contain only single elements, one obtains from Theorem 4.2 (*ii*) a triangular decomposition of Σ^{-1}, where $\urcorner$ is upper-triangular and $H = \Delta^{-1}$ is a diagonal matrix with all diagonal elements positive. This gives the parameters in a stepwise generating process for a covariance matrix which has been called a path analysis model by the geneticist Wright [43], [44], a system of linear recursive equations with uncorrelated residuals by the econometrician Wold [42], or, more recently, a linear triangular system [41].

For a mean-centered random column vector Y and ordering $(1, \ldots, d_N)$, such a process can be written in matrix notation as

$$AY = \varepsilon, \text{ with } \operatorname{cov}(\varepsilon) = \Delta. \tag{4.1}$$

having

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

In our notation, A is unit upper-triangular, so that from Theorem 4.2, and also $\urcorner$ and $\llcorner^{\mathrm{T}}$ of Illustration (3.1), the elements of a_{ij} of A and a^{ij} of A^{-1} are

$$a_{ij} = -\beta_{i|j.r(i)\setminus j}, \quad a^{ij} = \beta_{i|j.r(j)}, \tag{4.2}$$

where $\beta_{i|j.C}$ denotes the coefficient of Y_j in linear least squares regression of Y_i on Y_j and Y_C and C refers possibly to a vector variable, and where $r(k) = \{k+1, \ldots, d\}$. The diagonal elements of Δ are $\delta_{ii} = \sigma_{ii|r(i)}$, the residual variances in the corresponding linear regressions.

When $(1, \ldots, d)$ refers to a time order, then Y_1 is the most recent response variable and Y_d is the variable in the past, being most distant from it. The joint concentration matrix is directly generated by a sequence of univariate least squares regressions with Y_i as response to $Y_{i+1}, \ldots, Y_d$. Expressed differently, A contains the generating equation parameters and $(A, \Delta^{-1}$ gives the unique triangular decomposition of Σ^{-1} for one fixed order of the variables.

When there is zero contribution of a potentially explanatory variable Y_j for Y_i, this is represented by a zero value in position (i, j) in the upper triangular part of A. Variables with a nonzero contribution are called the parents of i, denoted by $\operatorname{par}(i)$. Equation (4.2) represents then an unconstrained model containing as a special case the reduced model [7] with

$$a_{ij} = -\beta_{i|j.\operatorname{par}(i)\setminus j} \text{ when } j \in \operatorname{par}(i), \quad 0 = \rho_{ij|\operatorname{par}(i)} = \rho_{ij|r(i)\setminus j} \text{ else.} \tag{4.3}$$

In general, availability of direct checks of goodness of fit of constraints is an important advantage of knowing a more general model for which explicit unconstrained estimates are available.

With $\mathrm{In}[M]$ denoting the indicator matrix of a matrix M, obtained by replacing every nonzero element of M by a one, the edge matrix of the independence graph of the generating process is $\mathcal{A} = \mathrm{In}[A]$ in which node i corresponds to variable Y_i. The graph is often called the parent graph since it shows the directly explanatory variables of Y_i by arrows starting in the parent node set $\mathrm{par}(i)$ and pointing to node i. It defines also the linear independence structure in a family of matrices in which the unconstrained equation parameters a_{ij} in equation (4.3) are free to vary, provided only that they lead to positive residual variances.

4.2 Relations to joint response models

Sequences of joint responses occur in different types of graphical chain models, used to study multivariate statistical dependence. These have been defined for more general than linear relations, but we discuss here only the linear case and introduce some more terminology first. All graphical chain models have in common that the variables are arranged in a sequence of say d chain components, named g, each containing one or more variables. Several variables in the same chain component are considered as joint responses, i.e. to be on equal footing, so that within chain components undirected associations are of interest and between chain components directed associations. Each missing edge in an associated chain graph captures a conditional independence constraint and each edge present a conditional association, the precise conditioning depends on the type of chain graph.

For $d = 1$, models with zero constraints on Σ^{-1} have been introduced as covariance selection models by Dempster [12]. Corresponding recursive sequences of such models for $1 < d < d_N$ are blocked-concentration chains defining ${}'\Sigma^{-1}$ [26], [17], [38], [35].

For $d = 1$, models with zero constraints on Σ have been introduced as hypotheses linear in covariances by Anderson [2] and studied later as independence models [24], [31]; we call corresponding recursive sequences of such models for $1 < d < d_N$ [23] blocked-covariance chains defining ${}'\Sigma$.

Models with zero constraints on $H = \Sigma_{gg|r(g)}$ and $\Pi_{g|r}$ are multivariate regression chains, which include seemingly unrelated regressions [8],[9], [31]. We speak of concentration-regression chains for models with zero constraints on elements of either component of the block-triangular decomposition $(H, \urcorner)$ of Σ^{-1} [3], while the partial regression chains with zero constraints on the block-triangular decomposition $(K, \mathsf{L}^{\mathrm{T}})$ of Σ appear to not have been studied by statisticians.

Matrix relations between parameters in linear chain graph models to those of a generating linear triangular system [41] are not repeated here, instead, the relations to partial inversion are spelled out.

Corollary 4.1. *Induced joint response models related to partial inversion.* Let $(1, \dots, g, \dots, d)$ be an ordered partitioning of N, defining block decompositions $\Sigma = \mathsf{L}^{\mathrm{T}} K \mathsf{L}$ and $\Sigma^{-1} = \urcorner^{\mathrm{T}} H \urcorner$ of Theorem 4.2 and $'\Sigma^{-1} = H\urcorner$, $\Sigma' = K\mathsf{L}$, then unconstrained parameters induced in

(*i*) $'\Sigma^{-1}$, for a blocked concentration chain;
(*ii*) H and $\urcorner$ for a concentration-regression chain;
(*iii*) Σ', for a blocked covariance chain;
(*iv*) K and L^{T} for a partial regression chain;
(*iv*) K and $\urcorner$, for a multivariate regression chain,

are in one-to-one correspondence to unconstrained specific parameters A, Δ of a given saturated triangular system and are obtainable by partial inversion.

Proof. It is implied by the partial inversion results of Theorem 4.2 and Corollary 3.2 that each of the sets of parameters is obtained by a one-to-one transformation of the covariance matrix. □

One important consequence of Corollary 4.1 for statistical analysis is, that unconstrained parameters, estimated by equating observed to expected moments, are in one-to-one correspondence for all these the different types of saturated models. Also, from the interpretation of the parameters in each of these model types, obtained with equation (3.4), we know the corresponding conditional independence statement when a parameter is constrained to be zero for a joint Gaussian distribution. And, more importantly, this provides the interpretation of each missing edge in all of the different types of chain graphs, in general.

4.3 Edge matrices induced after partial inversion for linear triangular systems

For the study of when and how independence constraints of a given stepwise generating process are preserved after partial inversion, we introduce for unit binary matrices $\mathcal{M}$ a structural zero operator written analogously to (2.2) as

$$\text{(4.4)}\quad \mathrm{zer}_a \mathcal{M} = \begin{pmatrix} \underline{\mathcal{M}}_{aa}^{-1} & \mathcal{M}_{a \leftharpoondown b} \\ \mathcal{M}_{b \rightharpoondown a} & \mathcal{M}_{bb.a} \end{pmatrix} = \mathrm{In} \begin{pmatrix} (\mathcal{M}_{aa}^{*})^{-1} & \underline{\mathcal{M}}_{aa}^{-1} \mathcal{M}_{ab} \\ \mathcal{M}_{ba} \underline{\mathcal{M}}_{aa}^{-1} & \mathcal{M}_{bb} + \mathcal{M}_{ba} \underline{\mathcal{M}}_{aa}^{-1} \mathcal{M}_{ab} \end{pmatrix}$$

where $\mathcal{M}_{aa}^{*}$ is an invertible matrix chosen such that $\mathrm{In}[\mathcal{M}_{aa}^{*}] = \mathcal{M}_{aa}$ and $(\mathcal{M}_{aa}^{*})^{-1} \geq 0$. We conjecture that it is always possible to find such a matrix $\mathcal{M}^{*}$ and prove existence for the two cases of interest for graphical chain models, that is for $\mathcal{M}$ symmetric and for $\mathcal{M}$ non-symmetric, but a permuted, unit upper-triangular matrix.

When $\mathcal{M} = \mathcal{A}$ is upper-triangular, it is the edge matrix of a parent graph, which is directed and acyclic and has a full ordering of the nodes. When $\mathcal{M} = \mathcal{S}$ is symmetric, it is the edge matrix of an undirected graph. Corresponding invertible matrices $\mathcal{A}^{*}$ and $\mathcal{S}^{*}$ with an unchanged zero structure and exclusively non-negative elements in their inverses are

$$\text{(4.5)}\qquad \mathcal{A}^{*} = 2\mathcal{I} - \mathcal{A}, \quad \mathcal{S}^{*} = (d+1)\mathcal{I} - \mathcal{S},$$

where $\mathcal{I}$ denotes the identity matrix. The binary matrices $\mathrm{In}[\mathcal{A}^*]$ and $\mathrm{In}[\mathcal{S}^*]$ are edge matrices of what is sometimes called the transitive closure of the graphs with edge matrices $\mathcal{A}$ and $\mathcal{S}$, respectively.

PROOF. The matrix $\mathcal{A}^*$ is the matrix form [32], p. 97, of Neumann's limit of a geometric power series [29], p. 29, and power r of $(\mathcal{A}-\mathcal{I})$ counts for each pair its number of direction-preserving paths in $r+1$ nodes. To prove the claimed property of $\mathcal{S}^*$, note that for a single undirected path the edge matrix $S_\sim$ is a tri-diagonal matrix of ones and the eigenvalues and eigenvectors of $S^*_\sim$ are known in explicit form. Then $(S^*_\sim)^{-1}$ multiplied by the determinant of $S^*_\sim$ is a matrix of positive integers with the smallest element in position $(1,d)$ equal to one. With additional ones added, the inverse of $\mathcal{S}^*$ in equation (4.5) remains nonnegative, the explicit form for the case without any zeros being again known in explicit form. For a graph consisting of several disconnected subgraphs, the edge matrix is of incomplete block-diagonal form and the same type of argument applies to the edge matrices of each connected subgraph. The matrix $\mathcal{S}^*$ is then complete block-diagonal and non-negative. □

The structural zero operator in equation (4.4) shares the commutativity (i) and the exchangeability (ii) properties with partial inversion when it defines partial closing of paths in graphs, i.e when $\mathcal{M}=\mathcal{S}$ or $\mathcal{M}=\mathcal{A}$. But, the operation cannot be undone when it is reapplied to the same rows and columns. The reason is that it is defined in terms of sums and products of nonnegative matrices so that zeros present in $\mathcal{M}$ may be preserved or removed but no new zeros can be generated. We conjecture that the following properties hold for general unit binary matrices as well.

THEOREM 4.2. *Commutativity, exchangeability and contraction/expansion for structural zeros preserved after partial inversion.* Let arbitrary components a,b,c,d partition N, $G=\{a,b\}$, and the matrix $\mathcal{M}$ be accordingly partitioned, then

$$\begin{aligned}
&(i)\quad \mathrm{zer}_a\,\mathrm{zer}_b\mathcal{M}=\mathrm{zer}_b\,\mathrm{zer}_a\mathcal{M}=\mathrm{zer}_{ab}\mathcal{M};\\
&(ii)\quad [\mathrm{zer}_a\mathcal{M}]_{G,G}=\mathrm{zer}_a\mathcal{M}_{GG};\\
&(iii)\ \mathrm{zer}_{ab}\,\mathrm{zer}_{bc}\mathcal{M}=\mathrm{zer}_{abc}\mathcal{M}\,.
\end{aligned}$$

Whenever a linear joint response chain is generated by a linear triangular system with independence structure, then the variable pairs for which independence constraints are preserved in the derived chain graph can be obtained by applying the structural zero operator. Consider any chain component, for which the relation of response Y_α on Y_β given Y_C is specified, and some variables Y_M are omitted, then $N=(M,\alpha,\beta,C)$ becomes the relevant ordering of the node set. For each chain component the sets to be marginalized over, M, and to be conditioned on, C, are to be redefined as well as $m=\{M,\alpha\}$, $c=\{C,\beta\}$ and $\mathcal{B}=\mathrm{zer}_m\mathcal{A}$. The derived edge matrix components [41] for $\mathrm{inv}_\alpha\mathrm{con}(Y_{\alpha|C},Y_{\beta|C})$

can then be written as

$$\begin{aligned}
\mathcal{S}_{\alpha\alpha|\beta C} &= \mathrm{In}[\mathcal{B}_{mm}\mathrm{zer}_m(\mathcal{I}_{mm} + \mathcal{B}_{cm}^{\mathrm{T}}\mathcal{B}_{cm})\,\mathcal{B}_{mm}^{\mathrm{T}}]_{\alpha,\alpha}\,; \\
\mathcal{S}^{\beta\beta.\alpha M} &= \mathrm{In}[\mathcal{B}_{cc}^{T}\,\mathrm{zer}_c(\mathcal{I}_{cc} + \mathcal{B}_{cm}\mathcal{B}_{cm}^{\mathrm{T}})\mathcal{B}_{cc}]_{\beta,\beta}\,; \\
\mathcal{P}_{\alpha|\beta.C} &= \mathrm{In}[\mathcal{B}_{mc} + \mathcal{B}_{mm}\mathcal{B}_{cm}^{\mathrm{T}}\mathrm{zer}_c(\mathcal{I}_{cc} + \mathcal{B}_{cm}\mathcal{B}_{cm}^{\mathrm{T}})\mathcal{B}_{cc}]_{\alpha,\beta}\,.
\end{aligned} \tag{4.6}$$

5 Discussion

The statistical importance of the derived or induced graphs of linear joint response chains defined with equations (4.6) is that they apply not only to Gaussian distributions but also to distributions of arbitrary form, provided that they are associated with the same parent graph [41]. This means that they satisfy all independencies specified by the given parent graph and that these independencies may be combined as in a non-degenerate Gaussian distribution [34].

Markov equivalence of two chain graphs means that the missing edges in the two graphs lead to the same set of independencies, i.e. specify an identical independence structure. For different special subclasses of chain graphs, criteria and algorithms to decide on Markov equivalence have been derived. For instance, a concentration graph model is Markov equivalent to a triangular system if and only if the concentration graph is chordal, i.e. it contains no chordless cycle in 4 or more nodes. Criteria for chordality of graphs have been given early, see e.g. [13], but efficient algorithms applicable to large graphs have been derived much later, see e.g. [27], [30].

Similarly, the explicit forms of edge matrices, given in equations (4.6) for joint response chain graphs generated by a given parent graph, need to be complemented by computationally efficient algorithms for applications to large matrices. A general computational problem is to start with the zero structure in any given triangular decomposition of Σ^{-1} and to obtain the implied structural zeros in a new block-triangular decomposition of Σ^{-1} or of Σ, i.e. to find the zeros that would be retained by corresponding symbolic matrix inversion. The problem is considerably more complex than checking for chordality of a concentration graph because of the blocking structure and because different components of block-triangular decompositions may require different types of edge.

Acknowledgement

We thank the Swedish Research Society for support of the cooperation of D.R. Cox and N. Wermuth, as well as the German Research Society and the University of Mainz for research grants. We also thank Søren Johansen, Ingram Olkin and the referees for their thoughtful and constructive comments.

REFERENCES

1. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, NY, 1958.

2. T. W. Anderson, *Asymptotically efficient estimation of covariance matrices with linear structure*, Ann. Statist., 1 (1973), pp. 135–141.
3. S. A. Andersson, D. Madigan, and M. D. Perlman, *Alternative Markov properties for chain graphs*, Scand. J. Statist, 28 (2001), pp. 33–86.
4. A. E. Beaton, *The use of special matrix operators in statistical calculus*, EdD. thesis, Harvard University, reprinted as Educational Testing Service Research Bulletin 64-51, Princeton, NJ, 1964.
5. A. L. Cholesky, *Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés à un sytème d'équations linéaires en nombre inférieur à celui des inconnues*, (Published six years after Cholesky's death by Benoit), Bulletin géodésique, 2 (1924), pp. 67–77.
6. W. G. Cochran, *The omission or addition of an independent variate in multiple linear regression*, Suppl. J. Roy. Statist. Soc., 5 (1938), pp. 171–176.
7. D. R. Cox and N. Wermuth, *An approximation to maximum-likelihood estimates in reduced models*, Biometrika, 77 (1990), pp. 747–761.
8. D. R. Cox and N. Wermuth, *Linear dependencies represented by chain graphs (with discussion)*, Statistical Science, 8 (1993), pp. 204–218; 247–277.
9. D. R. Cox and N. Wermuth, *Multivariate Dependencies: Models, Analysis, and Interpretation*, Chapman and Hall, London, 1996.
10. H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ, 1946.
11. A. P. Dempster, *Elements of Continuous Multivariate Analysis*, Addison-Wesley, Reading, MA, 1969.
12. A. P. Dempster, *Covariance selection*, Biometrics, 28 (1972), pp. 157–175.
13. G. A. Dirac, *On rigid circuit graphs*, Abhandlungen, Mathematisches Seminar Hamburg, 25 (1961), pp. 71–76.
14. M. H. Doolittle, *Explanations and illustrations of method employed in the office in solution of normal equations and the adjustment of a triangulation*, U.S. Coast and Geodetic Survey, Appendix 8 (1878), pp. 115–120.
15. D. Edwards, *Introduction to Graphical Modelling,* 2nd ed., Springer, New York, NY, 2000.
16. L. Fox, H. D. Huskey, and J. H. Wilkinson, *Notes on the solutions of algebraic linear simultaneous equations*, Quart. J. Mech. Appl. Math., 1 (1947), pp. 149–173.
17. M. Frydenberg, *The chain graph Markov property*, Scand. J. Statist., 17 (1990), pp. 333–353.
18. W. Gibbs, *Elementary Principles of Statistical Mechanics*, Yale University Press, New Haven, CT, 1902.
19. G. H. Golub and J. Y. Yuan, *Symmetric-triangular decomposition and its applications Part I: Theorems and algorithms*, BIT, 42 (2002), pp. 814–822.
20. J. P. Gram, *Über die Entwickling reeller Funktionen in Reihen mittelst der Methode der kleinsten Quadrate*, Crelle's Journal (Journal für die reine und angewandte Mathematik), 94 (1883), pp. 41–73.
21. W. Jordan, O. Eggert, and M. Kneissl, *Feld- und Landmessung*, Handbuch der Vermessungskunde, Metzler, Stuttgart, 1888.
22. S. Kangshen, J. N. Crossley, and A. W.-C. Lun, *The Nine Chapters on the Mathematical Art, Companion and Commentary*, English translation from Chinese, Oxford University Press, Oxford, 2000.

23. G. Kauermann, *On the analysis of independencies in margins of a multivariate distribution*, Dissertation, Department of Economics, Technical University Berlin, 1994.
24. G. Kauermann, *On a dualization of graphical Gaussian models*, Scand. J. Statist., 23 (1996), pp. 105–116.
25. S. L. Lauritzen, *Graphical Models*, Oxford University Press, Oxford, 1996.
26. S. L. Lauritzen and N. Wermuth, *Graphical models for associations between variables, some of which are qualitative and some quantitative*, *Ann. Statist.* 17 (1989), pp. 31–54.
27. H. G. Leimer, *Optimal decomposition by clique separators*, Discrete Mathematics, 113 (1993) pp. 99–123.
28. A. A. Markov, *Wahrscheinlichkeitsrechnung*, German translation of 2nd Russian edition in 1908, Teubner, Leipzig, 1912.
29. C. G. Neumann, *Vorlesungen über Riemann'sche Theorie der Abel'schen Integrale*, 2nd ed., Teubner, Leipzig, 1884.
30. B. W. Peyton, A. Pothen, and X. Yuan, *Partitioning a chordal graph into transitive subgraphs for parallel sparse triangular solution*, Linear Algebra Appl., 192 (1993) pp. 329–354.
31. T. S. Richardson, *Markov properties for acyclic directed mixed graphs*, Scand. J. Statist., 30 (2003), pp. 145–157.
32. S. R. Searle, *Matrix Algebra for the Biological Sciences*, Wiley, New York, NY, 1966.
33. E. Schmidt, *Entwicklungen willkürlicher Funktionen*, Mathematische Annalen, 63 (1907), pp. 433-476.
34. J. Q. Smith, *Influence diagrams for statistical modelling*, Ann. Statist., 17, (1989), pp. 654–672.
35. M. Studený and R. Bouckaert, *On chain graphs models for description of conditional independence structures*, Ann. Statist., 26 (1998), 1434-1495.
36. A. M. Turing, *Rounding-off errors in matrix processes*, Quart. J. Mech. Appl. Math, 1 (1948), pp. 287–308.
37. K. Vogel, *Neun Bücher arithmetischer Technik*, German translation from Chinese. Viehweg, Braunschweig, 1968.
38. N. Wermuth, *On block-recursive linear regression equations (with discussion)*, Brazilian Journal of Probability and Statistics (REBRAPE, Revista Braslileira de Probilidade e Estatistica), 6 (1992), pp. 1–56.
39. N. Wermuth, *Graphical chain models*, in Encyclopedia of Behavioral Statistics II, B. Everitt and David C. Howell, eds., Wiley, Chichester, 2005, pp. 755–757.
40. N. Wermuth and D. R. Cox, *On association models defined over independence graphs*, Bernoulli, 4 (1998), 477–494.
41. N. Wermuth and D. R. Cox, *Joint response graphs and separation induced by triangular systems*, J. Roy. Statist. Soc. B, 66 (2004), pp. 687–717.
42. H.O. Wold, *Causality and econometrics*, Econometrica, 22 (1954), pp. 162-177.
43. S. Wright, *The theory of path coefficients: a reply to Niles' Criticism*, Genetics, 8 (1923), pp. 239–255.
44. S. Wright, *The method of path coefficients*, Ann. Math. Statist., 5 (1934), pp. 161–215.