The Annals of Statistics
1993, Vol. 21, No. 3, 1318–1358

LATTICE MODELS FOR CONDITIONAL INDEPENDENCE IN A MULTIVARIATE NORMAL DISTRIBUTION[1]

By Steen Arne Andersson[2] and Michael D. Perlman

Indiana University and University of Washington

The lattice conditional independence model $\mathbf{N}(\mathscr{K})$ is defined to be the set of all normal distributions on $\mathbb{R}^I$ such that for every pair $L, M \in \mathscr{K}$, x_L and x_M are conditionally independent given $x_{L \cap M}$. Here $\mathscr{K}$ is a ring of subsets of the finite index set I and, for $K \in \mathscr{K}$, x_K is the coordinate projection of $x \in \mathbb{R}^I$ onto $\mathbb{R}^K$. Statistical properties of $\mathbf{N}(\mathscr{K})$ may be studied, for example, maximum likelihood inference, invariance and the problem of testing H_0: $\mathbf{N}(\mathscr{K})$ vs. H: $\mathbf{N}(\mathscr{M})$ when $\mathscr{M}$ is a subring of $\mathscr{K}$. The set $J(\mathscr{K})$ of join-irreducible elements of $\mathscr{K}$ plays a central role in the analysis of $\mathbf{N}(\mathscr{K})$. This class of statistical models occurs in the analysis of nonnested multivariate missing data patterns and nonnested dependent linear regression models.

1. Introduction. In this paper we define and study a class of conditional independence (CI) models determined by finite distributive lattices. For multivariate normal distributions, the parameter space and the likelihood function (LF) for such a lattice CI model can be factored into the products of parameter spaces and conditional LFs, respectively, corresponding to ordinary multivariate normal linear regression models. This in turn yields explicit maximum likelihood estimators (MLEs) and likelihood ratio tests (LRTs) by means of standard techniques from multivariate analysis.

These lattice CI models arise in a natural way in the analysis of multivariate missing data sets with nonmonotone missing data patterns [cf. Andersson and Perlman (1991)] and in the analysis of nonnested dependent linear regression models [cf. Andersson and Perlman (1993a)]. The factorizations mentioned previously can be readily applied to obtain explicit MLEs and LRTs by standard linear methods.

We introduce this class of lattice CI models by means of the following simple and familiar example. Let $(x_1, x_2, x_3)^t$ denote a random observation from the trivariate normal distribution $N(\Sigma)$ with mean vector 0 and unknown covariance matrix Σ. [For simplicity, throughout this paper we shall assume that Σ is nonsingular and that the mean vector of the sampled normal population is

Received February 1989; revised August 1992.

[1]Supported in part by the Danish Research Council, by NSF Grants DMS 86-03489 and 89-02211 and by U.S. National Security Agency Grant MDA 904-92-H-3083.

[2]Research carried out in part at the Institute of Mathematical Statistics, University of Copenhagen.

AMS 1991 *subject classifications*. Primary 62H12, 62H15; secondary 62H20, 62H25.

Key words and phrases. Distributive lattice, join-irreducible elements, pairwise conditional independence, multivariate normal distribution, generalized block-triangular matrices, maximum likelihood estimator, quotient space, nonnested missing data, nonnested linear regressions.

known. The latter assumption is easily removed; cf. Andersson and Perlman (1991, 1993a).] Consider the model that specifies that x_2 *and* x_3 *are conditionally independent given* x_1, which we express in the familiar notation

$$x_2 \perp\!\!\!\perp x_3 | x_1. \tag{1.1}$$

In terms of the covariance matrix Σ, (1.1) is equivalent to the condition

$$(\Sigma^{-1})_{23} = (\Sigma^{-1})_{32} = 0. \tag{1.2}$$

In order to express this as a lattice CI model, let $I \equiv \{1, 2, 3\}$ denote the index set and consider

$$\mathscr{K} \equiv \{\varnothing, \{1\}, \{1,2\}, \{1,3\}, I\}, \tag{1.3}$$

a subring of the ring $\mathscr{D}(I)$ of all subsets of I. Clearly, $\mathscr{K}$ is a finite distributive lattice under the usual set operations $\cup$ and $\cap$. Define the class $\mathbf{P}(\mathscr{K})$ of real positive definite $I \times I$ matrices as follows:

$$\Sigma \in \mathbf{P}(\mathscr{K}) \Leftrightarrow x_L \perp\!\!\!\perp x_M | x_{L \cap M} \qquad \forall\, L, M \in \mathscr{K}, \tag{1.4}$$

where $x \sim N(\Sigma)$ and x_T denotes the T-subvector of x when $T \subseteq I$. It is readily verified that (1.1), (1.2) and (1.4) are equivalent conditions. [Note that $x_2 \perp\!\!\!\perp x_3 | x_1 \Leftrightarrow (x_1, x_2) \perp\!\!\!\perp (x_1, x_3) | x_1$.]

In this example the factorizations of the parameter space and LF mentioned previously are represented as follows:

$$\Sigma \leftrightarrow \left(\Sigma_{11}, \Sigma_{21}\Sigma_{11}^{-1}, \Sigma_{22\cdot 1}, \Sigma_{31}\Sigma_{11}^{-1}, \Sigma_{33\cdot 1}\right), \tag{1.5}$$

$$f(x_1, x_2, x_3) = f(x_1)\, f(x_2|x_1)\, f(x_3|x_1). \tag{1.6}$$

The five parameters on the right-hand side of (1.5) represent ordinary unconditional and conditional variances and regression coefficients. Whereas the range of the positive definite matrix Σ in (1.5) is *constrained* by (1.2), the ranges of these five parameters are *unconstrained* (except for the trivial requirement that Σ_{11}, $\Sigma_{22\cdot 1}$ and $\Sigma_{33\cdot 1}$ are positive definite). Thus the MLEs of these five parameters, called the *$\mathscr{K}$-parameters* of the CI model, are easily obtained from (1.6), and the MLE of Σ may be reconstructed from these estimates.

A subset $K \in \mathscr{K}$ is called *join-irreducible* if K is not the join ($\equiv$ union) of two or more proper subsets of K (cf. subsection 2.1). The collection of all join-irreducible elements in $\mathscr{K}$ is denoted by $J(\mathscr{K})$. Thus, when $\mathscr{K}$ is given by (1.3),

$$J(\mathscr{K}) = \{\{1\}, \{1,2\}, \{1,3\}\}. \tag{1.7}$$

It will be seen that the basic factorizations (1.5) and (1.6), as well as their extensions to the general lattice CI model $\mathbf{N}(\mathscr{K})$ defined next, are always indexed by the members of $J(\mathscr{K})$.

Condition (1.4) immediately extends to define the general lattice CI model. Let I be an arbitrary finite index set and let $\mathscr{K}$ be an arbitrary subring of $\mathscr{D}(I)$, so again $\mathscr{K}$ is a finite distributive lattice. (By Birkhoff's theorem, *any* finite distributive lattice can be represented as a ring of subsets of some finite set I.) *Then* (1.4) *defines the class* $\mathbf{P}(\mathscr{K})$ *of* $I \times I$ *covariance matrices determined by CI restrictions with respect to the lattice* $\mathscr{K}$: $\Sigma \in \mathbf{P}(\mathscr{K})$ *if and only if* x_L *and* x_M *are conditionally independent given* $x_{L \cap M}$ *for every pair* $L, M \in \mathscr{K}$. If $N(\Sigma)$ denotes the normal distribution on $\mathbb{R}^I$ with mean vector 0 and unknown covariance matrix Σ, the normal statistical model

$$\mathbf{N}(\mathscr{K}) \equiv \{N(\Sigma) | \Sigma \in \mathbf{P}(\mathscr{K})\} \tag{1.8}$$

is *the lattice conditional independence* (*CI*) *model determined by* $\mathscr{K}$.

In this paper we study the structure of $\mathbf{P}(\mathscr{K})$ and the statistical properties of the model $\mathbf{N}(\mathscr{K})$. In Section 2.3 (Theorem 2.1) we generalize (1.2) by characterizing $\Sigma \in \mathbf{P}(\mathscr{K})$ in terms of the precision matrix Σ^{-1}. In Section 2.5 (Theorem 2.2) we generalize (1.5) by showing that each $\Sigma \in \mathbf{P}(\mathscr{K})$ can be uniquely represented in terms of its $\mathscr{K}$-parameters, whose ranges are unconstrained, so that the parameter space $\mathbf{P}(\mathscr{K})$ again factors into a product of parameter spaces for ordinary linear regression models. In Section 2.7 we present a general algorithm for reconstructing $\Sigma \in \mathbf{P}(\mathscr{K})$ from its $\mathscr{K}$-parameters. A series of examples in Section 2.8 illustrates these results.

The factorization (1.6) of the LF as a product of conditional densities involving only the $\mathscr{K}$-parameters of Σ is extended to the general lattice CI model $\mathbf{N}(\mathscr{K})$ in Section 3.1 (Theorem 3.1). The MLEs of the $\mathscr{K}$-parameters of Σ are easily derived from the general factorization, then the MLE of Σ can be reconstructed by the algorithm given in Section 2.7. This estimation procedure is illustrated by examples in Section 3.2. In Remark 3.5 it is noted that the model $\mathbf{N}(\mathscr{K})$ is determined by a system of *linear recursive equations* [cf. Wermuth (1980) or Kiiveri, Speed and Carlin (1984)] with additional lattice structure.

In Andersson and Perlman (1993b) we treat the problem of testing one lattice CI model against another, that is, testing

$$H_{\mathscr{K}}: \mathbf{N}(\mathscr{K}) \quad \text{vs.} \quad H_{\mathscr{M}}: \mathbf{N}(\mathscr{M}), \tag{1.9}$$

when $\mathscr{M}$ is a proper subring of $\mathscr{K}$. [Note that $\mathscr{M} \subset \mathscr{K} \Rightarrow \mathbf{N}(\mathscr{K}) \subseteq \mathbf{N}(\mathscr{M})$.] For example, in the simple case considered previously with $I = \{1, 2, 3\}$, suppose that $\mathscr{K} = \{\varnothing, \{1\}, \{1, 2\}, \{1, 3\}, I\}$ [cf. (1.3)] and $\mathscr{M} = \{\varnothing, I\}$. Then $\mathbf{N}(\mathscr{M})$ is simply the normal model with no restriction on Σ and (1.9) becomes the problem of testing $x_2 \perp\!\!\!\perp x_3 | x_1$ [equivalently, (1.2)] against the unrestricted alternative, which can be stated equivalently as the problem of testing

$$H_{\mathscr{K}}: \sigma_{23} = \sigma_{21}\sigma_{11}^{-1}\sigma_{13} \quad \text{vs.} \quad H_{\mathscr{M}}: \sigma_{23} \neq \sigma_{21}\sigma_{11}^{-1}\sigma_{13}, \tag{1.10}$$

where $\Sigma = (\sigma_{ij} | i, j = 1, 2, 3)$. If, however,

$$\mathscr{K} = \{\varnothing, \{1\}, \{3\}, \{1, 2\}, \{1, 3\}, I\}, \tag{1.11}$$

while $\mathscr{M} = \{\varnothing, \{1\}, \{1, 2\}, \{1, 3\}, I\}$, then (1.9) becomes the problem of testing $(x_1, x_2) \perp\!\!\!\perp x_3$ against $x_2 \perp\!\!\!\perp x_3 | x_1$, which is equivalent to the problem of testing

$$H_{\mathscr{K}}: \sigma_{13} = \sigma_{23} = 0 \quad \text{vs.} \quad H_{\mathscr{M}}: \sigma_{23} = \sigma_{21}\sigma_{11}^{-1}\sigma_{13}. \tag{1.12}$$

In fact, each of the testing problems treated by Das Gupta (1977), Giri (1979), Banerjee and Giri (1980) and Marden (1981), including (1.12), is a special case of the general problem (1.9).

[When $\mathscr{K}$ is given by (1.11) and $\mathscr{K}' = \{\varnothing, \{3\}, \{1,2\}, I\}$, then $\mathbf{N}(\mathscr{K}) = \mathbf{N}(\mathscr{K}')$, so both lattices determine the same CI conditions, namely $(x_1, x_2) \perp\!\!\!\perp x_3$. Thus two different lattices may determine the same CI model.]

The LRT statistic λ for the general testing problem (1.9) will be derived in Andersson and Perlman (1993b) and is readily expressible in terms of the MLEs of the $\mathscr{K}$-parameters and $\mathscr{M}$-parameters of Σ. The central distribution of λ may be obtained in terms of its moments by means of the invariance of the testing problem. Additional examples of this testing problem are treated in Andersson and Perlman (1993b).

These and associated results are greatly facilitated by the fact that the model $\mathbf{N}(\mathscr{K})$ is invariant under a group $G \equiv \mathbf{GL}(\mathscr{K})$ that acts transitively on $\mathbf{P}(\mathscr{K})$. This group G is a subgroup of a group of nonsingular block-triangular $I \times I$ matrices. To illustrate this, return to the trivariate lattice CI model considered previously with $\mathscr{K}$ given by (1.3). It can be seen that the CI model given by (1.1) $\equiv$ (1.2) is invariant under all nonsingular linear transformations of the form

$$x \equiv \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \equiv Ax. \tag{1.13}$$

The collection of all such matrices A forms a subgroup of the group of all 3×3 nonsingular lower triangular matrices. It is also true, but not so easy to see, that G acts *transitively* on the class $\mathbf{P}(\mathscr{K})$ of all covariance matrices Σ that satisfy (1.1) $\equiv$ (1.2), that is, for any such Σ there exists $A \in G$ such that $\Sigma = AA^t$.

These facts, some of which were used by Das Gupta (1977), Giri (1979), Banerjee and Giri (1980) and Marden (1981) to study the distribution and optimality of invariant tests for problems such as (1.10) and (1.12), will be extended in the present paper to the general lattice CI model $\mathbf{N}(\mathscr{K})$. In Section 2.4 it will be shown how $\mathscr{K}$ determines the invariance group $\mathbf{GL}(\mathscr{K})$, a group of generalized block-triangular $I \times I$ matrices with lattice structure, while the transitive action of $\mathbf{GL}(\mathscr{K})$ on $\mathbf{P}(\mathscr{K})$ is demonstrated in Section 2.6 (Theorem 2.3), generalizing the well-known Choleski decomposition of an arbitrary positive definite matrix. The transitivity yields a factorization (Lemma 2.5) of the determinant of $\Sigma \in \mathbf{P}(\mathscr{K})$, a generalization of the well-known Schur formula $\det(\Sigma) = \det(\Sigma_{11})\det(\Sigma_{22\cdot 1})$.

As already seen for the trivariate example given previously, all statistical properties of the general lattice CI model $\mathbf{N}(\mathscr{K})$, including the definition of the $\mathscr{K}$-parameters of Σ, the factorizations of its parameter space and LF as products of those for linear regression models, the form of the MLE, the form of the LRT statistic and its central distribution, and the partitioning and location of zeros in the invariance matrix $A \in \mathbf{GL}(\mathscr{K})$, are determined by the fundamental structure of the lattice $\mathscr{K}$, in particular, by the associated poset

$J(\mathscr{K})$ of *join-irreducible elements* of $\mathscr{K}$ (cf. Section 2.1). As in the case of a balanced ANOVA design where the poset of join-irreducible elements of the lattice of subspaces determines the ANOVA table [cf. Andersson (1990)], for the lattice CI model $\mathbf{N}(\mathscr{K})$ the poset $J(\mathscr{K})$ determines the statistical analysis of the model.

The CI models $\mathbf{N}(\mathscr{K})$ play an important role in the analysis of nonmonotone missing data models. Under the assumption of multivariate normality, it is well known that a *monotone* missing data model with *unrestricted* covariance matrix Σ admits a complete and explicit likelihood analysis, remaining invariant under the appropriate group of block-triangular matrices (in the usual sense), which acts transitively on the unrestricted set of covariance matrices [cf. Eaton and Kariya (1983) and Andersson, Marden and Perlman (1994)]. If the missing data pattern is *nonmonotone*, however, then explicit analysis is not possible in general.

The relationship between lattice CI models and nonmonotone missing data patterns is developed fully in Andersson and Perlman (1991) but can be illustrated in terms of the trivariate example considered previously. Suppose that one attempts to observe a random sample from the trivariate normal distribution $N(\Sigma)$, where Σ is unknown and initially unrestricted, but that some of the observations are incomplete. For example, suppose that we have several complete vector observations of the form $(x_1, x_2, x_3)^t$ and also several incomplete observations of the forms $(x_1, x_2)^t$ and $(x_1, x_3)^t$. Then the missing data pattern (actually, the pattern of the *observed* data) is the set

$$\mathscr{S} := \{\{1,2\},\{1,3\},\{1,2,3\}\}, \tag{1.14}$$

that is, the collection of subsets of $I \equiv \{1, 2, 3\}$ corresponding to the subvectors actually observed. Because the missing data pattern $\mathscr{S}$ is nonmonotone, that is, not totally ordered by inclusion, the LF cannot be factored into a product of LFs of linear regression models and the MLE of Σ cannot be obtained explicitly. (In fact, for some nonmonotone missing data patterns with insufficiently many complete observations, Σ may not be estimable.) Instead, iterative estimation methods such as the EM algorithm must be used, possibly accompanied by difficulties with convergence or uniqueness of the estimates [cf. Little and Rubin (1987)].

An alternate approach, suggested by Rubin (1987) and developed in Andersson and Perlman (1991), is to restrict Σ by imposing the CI conditions of the lattice CI model $\mathbf{N}(\mathscr{K})$, where $\mathscr{K} \equiv \mathscr{K}(\mathscr{S})$ is the lattice generated by $\mathscr{S}$. With $\mathscr{S}$ given by (1.14) it is easy to see that $\mathscr{K}$ is given by (1.3), so the corresponding CI condition is given by (1.1). Under this condition the densities for the complete and incomplete observations factor as

$$\begin{aligned} f(x_1, x_2, x_3) &= f(x_1)\, f(x_2|x_1)\, f(x_3|x_1), \\ f(x_1, x_2) &= f(x_1)\, f(x_2|x_1), \\ f(x_1, x_3) &= f(x_1)\, f(x_3|x_1), \end{aligned} \tag{1.15}$$

so the overall LF is a product of LFs of only the three types $f(x_1)$, $f(x_2|x_1)$ and

$f(x_3|x_1)$, the latter two corresponding to simple linear regression models. Also, the overall parameter space is the product of the parameter spaces for these three LFs. Therefore, the similar terms may be combined and the MLE of Σ may be obtained by maximizing these three LFs separately, which involves only elementary calculations. Furthermore, under the CI restriction $\Sigma \in \mathbf{P}(\mathscr{K})$, this nonmonotone missing data model remains invariant under the group $\mathbf{GL}(\mathscr{K})$ of lower triangular matrices A in (1.13) and $\mathbf{GL}(\mathscr{K})$ acts transitively on $\mathbf{P}(\mathscr{K})$. Finally, the CI assumption may be tested by means of the LRT for (1.10) as discussed previously.

Whereas the determination of the appropriate CI conditions and the factorization (1.15) is transparent in this simple example, a general missing data pattern requires the lattice-theoretic approach developed in the present paper; see Andersson and Perlman (1991) for complete details. Thus the results in the present paper open the possibility of applying classical multivariate techniques to a class of missing data models much larger than the monotone class [and, similarly, to a large class of nonnested dependent linear regression models; see Andersson and Perlman (1993a)].

In Section 4 the CI models and results already described are recast in an invariant ($\equiv$ coordinate-free) formulation, rather than in the matrix (coordinate-wise) formulation just given. This is done for the following reason: A model which, when presented in matrix formulation, may not appear to be a lattice CI model according to the noninvariant definition given previously, may in fact belong to this class after an appropriate linear transformation. (Of course, this is by no means unique to the lattice CI models. For example, the general balanced ANOVA model must be described invariantly in terms of a collection of orthogonal subspaces, rather than by specifying the values of certain coordinates of the mean vector.)

This is readily illustrated in terms of the trivariate missing data example given in the paragraph containing (1.14). Rather than the missing data pattern described by (1.14), consider a missing data array that includes incomplete observations involving not only the coordinates of x but also one or more linear combinations of these coordinates. For example, suppose that we have several complete observations of the form $(x_1, x_2, x_3)^t$ and also several incomplete observations of the forms $(x_1, x_2)^t$ and $(x_1 + x_2, x_3)^t$. Although this does not directly fit into the framework of the coordinate-wise missing data models discussed previously and in Andersson and Perlman (1991), it is easy to transform it to such a framework by means of a nonsingular linear transformation $(y_1, y_2, y_3) = (x_1 + x_2, x_2, x_3)$. In terms of y_1, y_2 and y_3 the missing data pattern is now given precisely by (1.14); hence as before the associated lattice CI model imposes the assumption that $y_2 \perp\!\!\!\perp y_3|y_1$, that is, $x_2 \perp\!\!\!\perp x_3|x_1 + x_2$ (equivalently, $x_1 \perp\!\!\!\perp x_3|x_1 + x_2$).

The existence and form of an appropriate linear transformation from x to y (or equivalently, of an appropriate vector basis for the observation space) may not be so apparent in more complex missing data schemes with linear combinations present. The invariant formulation of a general lattice CI model, presented in Section 4, allows one to recognize and treat, without a preliminary transformation, a set of CI conditions such as $x_2 \perp\!\!\!\perp x_3 | x_1 + x_2$ in the same manner as the coordinate-wise lattice CI conditions in (1.4).

The invariant formulation is stated in terms of a lattice $\mathscr{Q}$ of quotient spaces Q of a real finite-dimensional vector space V. (All vector spaces and matrices considered in this paper are defined over the field of real numbers. See Section 4.1 for definitions, where it is noted that if $\mathscr{Q}$ is distributive, then it is finite.) For each $Q \in \mathscr{Q}$ let p_Q: $V \to Q$ denote the projection onto Q. Then the general lattice CI model $\mathbf{N}(\mathscr{Q})$ is defined in Section 4.2 to be *the set of all nonsingular normal distributions $N(\sigma)$ on V with mean* 0 *and covariance σ such that p_R and p_T are conditionally independent given $p_{R \wedge T}$ for every pair $R, T \in \mathscr{Q}$*. In Theorem 4.1 it is noted that $\mathbf{N}(\mathscr{Q})$ is nonempty if and only if $\mathscr{Q}$ is distributive.

To express our original coordinate-wise formulation of the lattice CI models in this invariant framework, set $V = \mathbb{R}^I$, identify each subset $K \subseteq I$ with the quotient space $\mathbb{R}^K$ and let p_K: $\mathbb{R}^I \to \mathbb{R}^K$ denote the usual coordinate projection mapping. Then the definition of the general lattice CI model in the preceding paragraph reduces to (1.4).

The basic decomposition theorem for a distributive lattice $\mathscr{Q}$ of quotient spaces (cf. Section A.1) states that the observation space V can be represented as a product of vector spaces indexed by the poset $J(\mathscr{Q})$ of join-irreducible elements in $\mathscr{Q}$ in such a way that for each $Q \in \mathscr{Q}$, the projection p_Q: $V \to Q$ becomes simply a canonical projection. By means of this representation we may choose a *$\mathscr{Q}$-adapted basis* for V (cf. Proposition 4.1). In Section 4.3 it is shown that in terms of the coordinate system determined by this basis, the CI model $\mathbf{N}(\mathscr{Q})$ can be expressed in the canonical coordinate-wise form (1.4) and the statistical analysis of the model may then proceed according to the coordinate-wise formulation.

The general problem of testing one lattice CI model against another is formulated invariantly as follows: Test $H_{\mathscr{Q}}$: $\mathbf{N}(\mathscr{Q})$ vs. $H_{\mathscr{T}}$: $\mathbf{N}(\mathscr{T})$, where $\mathscr{Q}$ and $\mathscr{T}$ are distributive lattices of quotient spaces of V such that $\mathscr{T} \subset \mathscr{Q}$. In Section 4.4 it is noted that one can choose a basis for V that is both $\mathscr{Q}$-adapted and $\mathscr{T}$-adapted, by means of which this testing problem can be reduced to the canonical coordinate-wise form (1.9).

In recent years the study of multivariate dependence models defined by CI conditions determined by directed or undirected graphs has received increasing attention; see Whittaker (1990) for a readable introduction to this area. In some of these graphical models the CI assumptions are equivalent to the occurrence of patterns of zeros in the precision matrix Σ^{-1} of a multivariate normal distribution; hence the models are *linear* in Σ^{-1}. It will be seen from Examples 2.6–2.8, however, that, unlike the special case (1.2), in general the lattice CI models introduced here are neither linear in Σ^{-1} nor Σ. Furthermore, the statistical motivation and analysis of a lattice CI model appear to differ from those of a model defined initially by graphical CI conditions.

2. The class $\mathbf{P}(\mathscr{K})$ of covariance matrices Σ determined by pairwise conditional independence with respect to a finite distributive lattice $\mathscr{K}$. Let I be a finite index set, let $\mathscr{D}(I)$ denote the ring of all subsets of I and let $\mathscr{K} \subseteq \mathscr{D}(I)$ be a subring, that is, $\mathscr{K}$ is closed under $\cap$ and $\cup$. We shall always assume that $I, \varnothing \in \mathscr{K}$. Then $\mathscr{K}$ is a finite distributive lattice with $\cup$ and $\cap$ as the join and meet operations, respectively. For $T, U \in \mathscr{D}(I)$ we write $T \subset U$ to indicate that $T \subseteq U$ but $T \neq U$. Let $|T|$ denote the number of elements in a set T.

Let $N(\Sigma)$ denote the normal distribution on $\mathbb{R}^I$ with mean $0 \in \mathbb{R}^I$ and covariance matrix $\Sigma \in \mathbf{P}(I)$, where $\mathbf{P}(I)$ denotes the set of all positive definite $I \times I$ matrices. For any $T \subseteq I$ and column vector $x = (x_i | i \in I) \in \mathbb{R}^I$, define $x_T := (x_i | i \in T)$, the T-subcolumn of x. Note that $x_I \equiv x$ and define $x_\varnothing := \{0\}$.

DEFINITION 2.1. The class $\mathbf{P}(\mathscr{K}) \subseteq \mathbf{P}(I)$ is defined [in Andersson and Perlman (1991), $\mathbf{P}(\mathscr{K})$ was denoted as $\mathbf{P}_{\mathscr{K}}(I)$] as follows [cf. (1.4)]:

$$(2.1) \quad \Sigma \in \mathbf{P}(\mathscr{K}) \Leftrightarrow x_L \perp\!\!\!\perp x_M | x_{L \cap M} \qquad \forall\, L, M \in \mathscr{K} \text{ when } x \sim N(\Sigma),$$

that is, x_L and x_M are conditionally independent (CI) given $x_{L \cap M}$ $\forall\, L, M \in \mathscr{K}$.

If $L \cap M = \varnothing$, then (2.1) reduces to $x_L \perp\!\!\!\perp x_M$, that is, x_L and x_M are independent. Note that the right-hand side of (2.1) is ordinarily written in the form

$$(2.2) \qquad x_{L \setminus (L \cap M)} \perp\!\!\!\perp x_{M \setminus (L \cap M)} | x_{L \cap M} \qquad \forall\, L, M \in \mathscr{K}.$$

Some of these pairwise CI conditions are trivially satisfied, for example, whenever $L \subseteq M$ (or $M \subseteq L$) (also see Remark 3.2). In particular, if $\mathscr{K}$ is a chain, then $\mathbf{P}(\mathscr{K}) = \mathbf{P}(I)$, that is, Σ is unrestricted (cf. Examples 2.1 and 2.2).

2.1. *The poset $J(\mathscr{K})$ of join-irreducible elements.* The structure of $\Sigma \in \mathbf{P}(\mathscr{K})$ will be characterized in terms of the poset $J(\mathscr{K})$ of *join-irreducible elements* of $\mathscr{K}$, which we now define. For $K \in \mathscr{K}$, $K \neq \varnothing$, define

$$\langle K \rangle := \bigcup (K' \in \mathscr{K} | K' \subset K),$$

$$[K] := K \setminus \langle K \rangle,$$

so that

$$(2.3) \qquad K = \langle K \rangle \,\dot{\cup}\, [K],$$

where $\dot{\cup}$ indicates that the union is disjoint. Then define

$$\begin{aligned} J(\mathscr{K}) &:= \{K \in \mathscr{K} | K \neq \varnothing, \langle K \rangle \subset K\} \\ &= \{K \in \mathscr{K} | K \neq \varnothing, [K] \neq \varnothing\} \\ &= \{K \in \mathscr{K} | K \neq \varnothing, \forall\, L, M \in \mathscr{K}\colon K = L \cup M \Rightarrow K = L \text{ or } K = M\}. \end{aligned}$$

If $K \in J(\mathscr{K})$ we say that K is *join-irreducible*. [See Grätzer (1978), Chapter II, or Davey and Priestley (1990), Chapter 8, for properties of $J(\mathscr{K})$; in particular, $\mathscr{K}$ is uniquely determined by $J(\mathscr{K})$.]

For $L \in \mathscr{K}$ define $\mathscr{K}_L := \{K \in \mathscr{K} | K \subseteq L\}$, a sublattice of $\mathscr{K}$ ($\mathscr{K}_I \equiv \mathscr{K}$). The following relations are elementary but fundamental:

$$L = \bigcup (K \in J(\mathscr{K}_L)), \tag{2.4}$$

$$J(\mathscr{K}_L) = J(\mathscr{K}) \cap \mathscr{K}_L, \tag{2.5}$$

$$J(\mathscr{K}_{L \cap M}) = J(\mathscr{K}_L) \cap J(\mathscr{K}_M), \tag{2.6}$$

$$J(\mathscr{K}_{L \cup M}) = J(\mathscr{K}_L) \cup J(\mathscr{K}_M). \tag{2.7}$$

PROPOSITION 2.1. *Every $L \in \mathscr{K}$ can be decomposed according to the members of $J(\mathscr{K})$ as follows*:

$$L = \dot{\bigcup} ([K] | K \in J(\mathscr{K}_L)). \tag{2.8}$$

PROOF. Let $K, M \in J(\mathscr{K})$ with $K \neq M$, so that $K \cap M \subset K$ or $K \cap M \subset M$. Suppose that $K \cap M \subset M$. Then $K \cap M \subseteq \langle M \rangle$ and it follows that $[K] \cap [M] \equiv K \cap \langle K \rangle^c \cap M \cap \langle M \rangle^c = \varnothing$; hence $([K] | K \in J(\mathscr{K}))$ is a disjoint family. The inclusion $\supseteq$ in (2.8) is trivial. To establish $\subseteq$, consider $\iota \in L$. Define $K_\iota =: \bigcap (L' \in \mathscr{K}_L | \iota \in L')$, the smallest set in $\mathscr{K}_L$ containing ι. Then $K_\iota \in J(\mathscr{K}_L)$, as seen from the following indirect argument. Suppose that $K_\iota \notin J(\mathscr{K}_L)$ and thus that $K_\iota = L_1 \cup L_2$, where $L_1, L_2 \in \mathscr{K}_L$, $L_1 \subset K_\iota$ and $L_2 \subset K_\iota$. Then $\iota \in L_1$ or $\iota \in L_2$, contradicting the minimality of K_ι. Finally, if $\iota \in \langle K_\iota \rangle$ ($\subset K_\iota$) the minimality of K_ι again would be contradicted; hence $\iota \in [K_\iota]$. Since $K_\iota \in J(\mathscr{K}_L)$ this establishes the inclusion $\subseteq$ in (2.8). □

In particular, set $L = I$ in (2.8) to obtain

$$I = \dot{\bigcup} ([K] | K \in J(\mathscr{K})). \tag{2.9}$$

For example, suppose that $I = \{1, 2, 3\}$ and $\mathscr{K}$ is given by (1.3). Then $J(\mathscr{K})$ is given by (1.7) and we find that $[\{1\}] = \{1\}$, $[\{1, 2\}] = \{2\}$ and $[\{1, 3\}] = \{3\}$, so (2.9) is evident.

2.2. *The $\mathscr{K}$-parameters of* Σ. For any finite index sets T and U, let $\mathbf{M}(T \times U)$ denote the vector space of all $T \times U$ matrices, $\mathbf{P}(T)$ the cone of all positive definite $T \times T$ matrices, $\mathbf{M}(T) \equiv \mathbf{M}(T \times T)$ the algebra of all $T \times T$ matrices and $\mathbf{GL}(T)$ the group of all nonsingular $T \times T$ matrices. For every $\Sigma \in \mathbf{P}(I)$ and every subset $T \subseteq I$, let Σ_T $[\in \mathbf{P}(T)]$ denote the $T \times T$ submatrix of Σ and let Σ_T^{-1} denote $(\Sigma_T)^{-1}$. For $K \in J(\mathscr{K})$ partition Σ_K according to (2.3) as follows:

$$\Sigma_K = \begin{pmatrix} \Sigma_{\langle K \rangle} & \Sigma_{\langle K]} \\ \Sigma_{[K \rangle} & \Sigma_{[K]} \end{pmatrix}, \tag{2.10}$$

so $\Sigma_{\langle K \rangle} \in \mathbf{P}(\langle K \rangle)$, $\Sigma_{[K]} \in \mathbf{P}([K])$, $\Sigma_{[K \rangle} \in \mathbf{M}([K] \times \langle K \rangle)$ and $\Sigma_{\langle K]} = (\Sigma_{[K \rangle})^t$.

Furthermore, define

$$\Sigma_{[K]\cdot} \equiv \Sigma_{[K]\cdot\langle K\rangle} := \Sigma_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}\Sigma_{\langle K]} \in \mathbf{P}([K]) \tag{2.11}$$

and let $\Sigma_{[K]\cdot}^{-1}$ denote $(\Sigma_{[K]\cdot})^{-1}$. Then for every $x \in \mathbb{R}^I$,

$$\begin{aligned} &\operatorname{tr}\left(\Sigma_K^{-1} x_K x_K^t\right) \\ &\quad = \operatorname{tr}\left(\Sigma_{[K]\cdot}^{-1}\left(x_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle}\right)(\cdots)^t\right) + \operatorname{tr}\left(\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle}x_{\langle K\rangle}^t\right). \end{aligned} \tag{2.12}$$

DEFINITION 2.2. For $\Sigma \in \mathbf{P}(I)$, the family of matrices

$$\left(\left(\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}, \Sigma_{[K]\cdot}\right) \middle| K \in J(\mathscr{K})\right) \tag{2.13}$$

is called the family of *$\mathscr{K}$-parameters* of Σ.

2.3. *Characterization of conditional independence in terms of* Σ^{-1}. Theorem 2.1 presents an algebraic characterization of the set $\mathbf{P}(\mathscr{K})$ of covariance matrices Σ defined in terms of pairwise conditional independence [cf. (2.1)]. The following description of pairwise CI is useful.

LEMMA 2.1. *Let* $x \sim N(\Sigma)$, $\Sigma \in \mathbf{P}(I)$. *Then for any* $L, M \subseteq I$, $x_L \perp\!\!\!\perp x_M | x_{L\cap M}$ *if and only if* $\forall\, x \in \mathbb{R}^I$:

$$\begin{aligned} \operatorname{tr}\left(\Sigma_{L\cup M}^{-1} x_{L\cup M} x_{L\cup M}^t\right) &= \operatorname{tr}\left(\Sigma_L^{-1} x_L x_L^t\right) + \operatorname{tr}\left(\Sigma_M^{-1} x_M x_M\right) \\ &\quad - \operatorname{tr}\left(\Sigma_{L\cap M}^{-1} x_{L\cap M} x_{L\cap M}^t\right). \end{aligned} \tag{2.14}$$

PROOF. The difference

$$\operatorname{tr}\left(\Sigma_{L\cup M}^{-1} x_{L\cup M} x_{L\cup M}^t\right) - \operatorname{tr}\left(\Sigma_{L\cap M}^{-1} x_{L\cap M} x_{L\cap M}^t\right)$$

appears in the exponential term of the conditional density of $x_{(L\cup M)\setminus(L\cap M)}$ given $x_{L\cap M}$. Therefore, $x_L \perp\!\!\!\perp x_M | x_{L\cap M}$ if and only if this difference is the sum of the differences appearing in the exponential terms of the conditional densities of $x_{L\setminus(L\cap M)}$ given $x_{L\cap M}$ and $x_{M\setminus(L\cap M)}$ given $x_{L\cap M}$. This sum is

$$\begin{aligned} &\left(\operatorname{tr}\left(\Sigma_L^{-1} x_L x_L^t\right) - \operatorname{tr}\left(\Sigma_{L\cap M}^{-1} x_{L\cap M} x_{L\cap M}^t\right)\right) \\ &\quad + \left(\operatorname{tr}\left(\Sigma_M^{-1} x_M x_M^t\right) - \operatorname{tr}\left(\Sigma_{L\cap M}^{-1} x_{L\cap M} x_{L\cap M}^t\right)\right), \end{aligned}$$

and the lemma follows. □

THEOREM 2.1 [Characterization of $\mathbf{P}(\mathscr{K})$]. *For* $\Sigma \in \mathbf{P}(I)$ *the following conditions are equivalent*:

(i) $\Sigma \in \mathbf{P}(\mathscr{K})$;

(ii) $\forall\, x \in \mathbb{R}^I$:

$$\operatorname{tr}(\Sigma^{-1} x x^t) = \sum\left(\operatorname{tr}\left(\Sigma_{[K]\cdot}^{-1}\left(x_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle}\right)(\cdots)^t\right) \middle| K \in J(\mathscr{K})\right);$$

(iii) $\forall\, x \in \mathbb{R}^I, \forall\, L \in \mathscr{K}$:

$$\operatorname{tr}(\Sigma_L^{-1} x_L x_L^t) = \sum\left(\operatorname{tr}\left(\Sigma_{[K]\cdot}^{-1}\left(x_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle}\right)(\cdots)^t\right) \middle| K \in J(\mathscr{K}_L)\right).$$

PROOF. To show (i) $\Rightarrow$ (ii), use induction on $|J(\mathscr{K})| =: q$. If $q = 1$, then by (2.4), $\mathscr{K} = \{\varnothing, I\}$ and (ii) is trivial. Next, assume that (i) $\Rightarrow$ (ii) whenever $q \leq k-1$ and suppose that $q = k$. If $I \in J(\mathscr{K})$, then $J(\mathscr{K}) = J(\mathscr{K}_{\langle I \rangle}) \dot{\cup} \{I\}$; hence $|J(\mathscr{K}_{\langle I \rangle})| = k-1$. By the induction assumption, (iii) is true with L replaced by $\langle I \rangle$, so (ii) follows from (2.12) with K replaced by I. If, on the other hand, $I \notin J(\mathscr{K})$, then $I = L \cup M$ where $L \subset I$ and $M \subset I$. It follows from (2.4) that $|J(\mathscr{K}_L)| < k$ and $|J(\mathscr{K}_M)| < k$, so by the induction assumption (iii) is valid with L replaced by L, M and $L \cap M$. Then (ii) follows from (2.6), (2.7) and Lemma 2.1.

To show (ii) $\Rightarrow$ (iii), let $p: \mathbb{R}^I \to \mathbb{R}^L$ be the coordinate projection and define the subspace $U \subseteq \mathbb{R}^I$ as follows:

$$U := \left\{x \in \mathbb{R}^I | x_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle} = 0, \quad K \in J(\mathscr{K}),\, K \not\subseteq L\right\}.$$

By (ii) and the Polarization Identity, for every $x, y \in \mathbb{R}^I$,

$$\operatorname{tr}(\Sigma^{-1}xy^t) = \sum\Big(\operatorname{tr}\Big(\Sigma_{[K]\cdot}^{-1}\big(x_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle}\big) \times \big(y_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}y_{\langle K\rangle}\big)^t\Big)\Big|K \in J(\mathscr{K})\Big).$$

For $x \in U$ those summands with $K \not\subseteq L$ vanish, while for $y \in p^{-1}(0)$ those summands with $K \subseteq L$ vanish, hence $U \subseteq (p^{-1}(0))^{\perp}$, the orthogonal complement of $p^{-1}(0)$ with respect to the inner product determined by (ii). But $x \in (p^{-1}(0))^{\perp}$ iff

$$\operatorname{tr}(\Sigma^{-1}xx^t) = \operatorname{tr}\big(\Sigma_L^{-1}x_L x_L^t\big),$$

while if $x \in U$ then (ii) implies

$$\operatorname{tr}(\Sigma^{-1}xx^t) = \sum\Big(\operatorname{tr}\Big(\Sigma_{[K]\cdot}^{-1}\big(x_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}x_{\langle K\rangle}\big)(\cdots)^t\Big)\Big|K \in J(\mathscr{K}),\, K \subseteq L\Big),$$

so (iii) must hold for every $x \in U$. Since (iii) depends on $x \in (\mathrm{R}^I$ only through $p(x) \equiv x_L$, in order to establish (iii) for every $x \in \mathbb{R}^I$ it suffices to show that $p(U) = \mathbb{R}^L$. Clearly $p(U) \subseteq \mathbb{R}^L$. Conversely, for any $v \in \mathbb{R}^L$ we can construct $x \in U$ such that $p(x) = v$, as follows. Set $x_L = v$. Let $K_1, \dots, K_q$ be a never-decreasing listing of the members of $J(\mathscr{K})$ (cf. Remark 2.1 and the notational conventions in Lemma 2.4, and Section 2.7). By (2.4) and (2.8), this listing may be chosen such that $L = \bigcup([K_j]|j = 1, \dots, r)$, where $1 \leq r \leq q$. Now set $x_{[k]} = \Sigma_{[k\rangle}\Sigma_{\langle k\rangle}^{-1}x_{\langle k\rangle}$ successively for $k = r+1, \dots, q$. By (2.42) and the first of the two relations that precede it, this process is well defined, while by (2.9) it completely determines $x := (x_L, x_{[r+1]}, \dots, x_{[q]}) \in \mathbb{R}^I$. Clearly $x \in U$ and $p(x) = v$.

To show (iii) $\Rightarrow$ (i), consider any pair $L, M \in \mathscr{K}$. Apply condition (iii) four times, with L replaced by $L \cup M$, L, M and $L \cap M$, and then apply (2.6) and (2.7) to obtain (2.14). By Lemma 2.1, therefore, (i) is satisfied. □

2.4. *The $\mathscr{K}$-preserving matrices: generalized block-triangular matrices with lattice structure.* We now introduce a group $\mathbf{GL}(\mathscr{K})$ of nonsingular matrices A that will be seen in Section 2.6 to act transitively on $\mathbf{P}(\mathscr{K})$. In the present section $\mathbf{GL}(\mathscr{K})$ is shown to be a group of block-triangular matrices with lattice structure determined by $\mathscr{K}$.

For $A \in \mathbf{M}(I)$, $K \in \mathscr{K}$ and $L, M \in J(\mathscr{K})$, let A_K (resp., $A_{[LM]}$) denote the $K \times K$ (resp., $[L] \times [M]$) submatrix of A.

PROPOSITION 2.2. *Let $A \in \mathbf{M}(I)$. The following three conditions on A are equivalent*:

$$\text{(i)} \qquad \forall\, x \in \mathbb{R}^I, \forall\, L \in \mathscr{K}: \quad x_L = 0 \Rightarrow (Ax)_L = 0;$$

$$\text{(ii)} \qquad \forall\, x \in \mathbb{R}^I, \forall\, L \in \mathscr{K}: \quad (Ax)_L = A_L x_L;$$

$$\text{(iii)} \qquad \forall\, L, M \in J(\mathscr{K}): \qquad M \not\subseteq L \Rightarrow A_{[LM]} = 0.$$

PROOF. (ii) $\Rightarrow$ (i) is trivial.

(iii) $\Rightarrow$ (ii). By the usual formula for matrix multiplication by blocks,

$$\begin{aligned}(Ax)_L &= \left(\sum \left(A_{[KM]} x_{[M]} | M \in J(\mathscr{K})\right) | K \in J(\mathscr{K}_L)\right)\\ &= \left(\sum \left(A_{[KM]} x_{[M]} | M \in J(\mathscr{K}_L)\right) | K \in J(\mathscr{K}_L)\right)\\ &= A_L x_L.\end{aligned}$$

The first equality uses (2.8) and (2.9), the second uses condition (iii), while the third uses (2.8) twice.

(i) $\Rightarrow$ (iii). Suppose $L, M \in J(\mathscr{K})$ with $M \not\subseteq L$. Let ε denote any column vector in $\mathbb{R}^I$ satisfying $\varepsilon_{[K]} = 0$ for $K \in J(\mathscr{K})$, $K \neq M$. Then

$$A_{[LM]} \varepsilon_{[M]} = \sum \left(A_{[LK]} \varepsilon_{[K]} | K \in J(\mathscr{K})\right) = (A\varepsilon)_{[L]}.$$

But $(A\varepsilon)_L = 0$ by (i); hence $(A\varepsilon)_{[L]} = 0$. Since $\varepsilon_{[M]}$ is arbitrary this implies $A_{[LM]} = 0$ as required. □

Let $\mathbf{M}(\mathscr{K})$ denote the set of all $A \in \mathbf{M}(I)$ that satisfy the equivalent conditions (i), (ii) and (iii) in Proposition 2.2 and let $\mathbf{GL}(\mathscr{K})$ denote the set of all nonsingular matrices in $\mathbf{M}(\mathscr{K})$. It follows from (i) that $\mathbf{M}(\mathscr{K})$ is a matrix algebra and hence $\mathbf{GL}(\mathscr{K})$ is a matrix group. It also follows by (i) that $\mathbf{M}(\mathscr{K})$ is the set of all matrices that, for each $L \in \mathscr{K}$, preserve the kernel of the projection $\mathbb{R}^I \to \mathbb{R}^L$ given by $x \to x_L$. Note that when $\mathscr{K} = \{\varnothing, I\}$, $\mathbf{M}(\mathscr{K}) = \mathbf{M}(I)$ and $\mathbf{GL}(\mathscr{K}) = \mathbf{GL}(I)$.

DEFINITION 2.3. The algebra $\mathbf{M}(\mathscr{K})$ is called *the algebra of $\mathscr{K}$-preserving matrices* and $\mathbf{GL}(\mathscr{K})$ *the group of $\mathscr{K}$-preserving matrices*.

REMARK 2.1. When $\mathscr{K}$ is a chain then $J(\mathscr{K}) \equiv \mathscr{K} \setminus \{\varnothing\}$ is also a chain, so it follows from Proposition 2.2(iii) that $\mathbf{M}(\mathscr{K})$ is an algebra of block-triangular matrices in the usual sense. For a general $\mathscr{K}$ let $q := |J(\mathscr{K})|$ and let $K_1, K_2, \dots, K_q$ be a *never-decreasing* listing of the members of the poset $J(\mathscr{K})$, that is, $i < j \Rightarrow K_j \not\subseteq K_i$. If every $A \in \mathbf{M}(I)$ is partitioned according to the *ordered* decomposition

$$I = [K_1]\,\dot{\cup}\,[K_2]\,\dot{\cup}\cdots\dot{\cup}\,[K_q]. \tag{2.15}$$

then it is seen from Proposition 2.2(iii) that $\mathbf{M}(\mathscr{K})$ can be represented as a subalgebra of the algebra of lower block-triangular matrices. That is, $A \in \mathbf{M}(\mathscr{K})$ is lower block-triangular with additional blocks of zeros below the main diagonal; see (1.13) and also Section 2.8 for further examples.

REMARK 2.2. For $K \in J(\mathscr{K})$ and $A \in \mathbf{M}(I)$, partition A_K according to (2.3) and (2.10) as follows:

$$A_K = \begin{pmatrix} A_{\langle K\rangle} & A_{\langle K]} \\ A_{[K\rangle} & A_{[K]} \end{pmatrix}; \tag{2.16}$$

note that $A_{[KK]} = A_{[K]}$ when $K \in J(\mathscr{K})$. By Proposition 2.2(ii), if $A \in \mathbf{M}(\mathscr{K})$ then for every $K \in J(\mathscr{K})$ and $x \in \mathbb{R}^I$,

$$A_{\langle K]} = 0, \tag{2.17}$$

$$(Ax)_{[K]} = A_{[K]}x_{[K]} + A_{[K\rangle}x_{\langle K\rangle}. \tag{2.18}$$

Furthermore, the linear mapping

$$\begin{aligned} \mathbf{M}(\mathscr{K}) &\to \times(\mathbf{M}([K]\times\langle K\rangle)\times\mathbf{M}([K])|K\in J(\mathscr{K})), \\ A &\to \big((A_{[K\rangle}, A_{[K]})|K\in J(\mathscr{K})\big) \end{aligned} \tag{2.19}$$

is bijective. This holds because, by Proposition 2.2(iii), $A \in \mathbf{M}(\mathscr{K})$ if and only if the $[K]\times(I\setminus K)$-submatrix of A is 0 for every $K \in J(\mathscr{K})$. Under the correspondence (2.19) the subset $\mathbf{GL}(\mathscr{K})$ corresponds to the subset

$$\times\,(\mathbf{M}([K]\times\langle K\rangle)\times\mathbf{GL}([K])|K\in J(\mathscr{K})). \tag{2.20}$$

LEMMA 2.2. *For $A \in \mathbf{GL}(\mathscr{K})$, $L \in \mathscr{K}$ and $K \in J(\mathscr{K})$,*

$$(A^{-1})_L = (A_L)^{-1} =: A_L^{-1}, \tag{2.21}$$

$$(A^{-1})_{[K]} = (A_{[K]})^{-1} =: A_{[K]}^{-1}, \tag{2.22}$$

$$(A^{-1})_{[K\rangle}A_{\langle K\rangle} = -A_{[K]}^{-1}A_{[K\rangle}. \tag{2.23}$$

PROOF. From Proposition 2.2(ii), $(AC)_L = A_L C_L$ for every $A, C \in \mathbf{M}(\mathscr{K})$, $L \in \mathscr{K}$, which implies (2.21). Then (2.22) and (2.23) follow from (2.17). □

LEMMA 2.3. *The mapping*

$$\begin{aligned} \mathbf{P}(I) &\to \times(\mathbf{M}([K]\times\langle K\rangle)\times\mathbf{P}([K])|K\in J(\mathscr{K})), \\ \Sigma &\to \big((\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}, \Sigma_{[K]\cdot})|K\in J(\mathscr{K})\big) \end{aligned} \tag{2.24}$$

from Σ to its $\mathscr{K}$-parameters commutes with the actions of $\mathbf{GL}(\mathscr{K})$ on $\mathbf{P}(I)$

and on $\times(\mathbf{M}([K]\times\langle K\rangle)\times\mathbf{P}([K])|K\in J(\mathscr{K}))$ *given by*

$$\mathbf{GL}(\mathscr{K})\times\mathbf{P}(I)\to\mathbf{P}(I),\qquad (A,\Sigma)\to A\Sigma A^t \tag{2.25}$$

and

$$\begin{aligned}&\mathbf{GL}(\mathscr{K})\times(\times(\mathbf{M}([K]\times\langle K\rangle)\times\mathbf{P}([K])|K\in J(\mathscr{K})))\\&\quad\to\times(\mathbf{M}([K]\times\langle K\rangle)\times\mathbf{P}([K])|K\in J(\mathscr{K})),\\&\big(A,((R_{[K\rangle},\Lambda_{[K]})|K\in J(\mathscr{K}))\big)\\&\quad\to\big((A_{[K]}R_{[K\rangle}A^{-1}_{\langle K\rangle}+A_{[K\rangle}A^{-1}_{\langle K\rangle},A_{[K]}\Lambda_{[K]}A^t_{[K]})|K\in J(\mathscr{K})\big),\end{aligned}\tag{2.26}$$

respectively.

PROOF. It is straightforward to verify that (2.26) is a group action. We must show that for every $A\in\mathbf{GL}(\mathscr{K})$, $\Sigma\in\mathbf{P}(I)$ and $K\in J(\mathscr{K})$,

$$(A\Sigma A^t)_{[K\rangle}(A\Sigma A^t)^{-1}_{\langle K\rangle}=A_{[K]}\Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}A^{-1}_{\langle K\rangle}+A_{[K\rangle}A^{-1}_{\langle K\rangle} \tag{2.27}$$

and

$$(A\Sigma A^t)_{[K]\cdot}=A_{[K]}\Sigma_{[K]\cdot}A^t_{[K]}. \tag{2.28}$$

It follows from Proposition 2.2(ii) that $(A\Sigma A^t)_K=A_K\Sigma_K A^t_K$. Let A_K and Σ_K be partitioned as in (2.16) and (2.10), respectively. Since $A_{\langle K]}=0$, (2.27) and (2.28) follow by direction calculation. □

PROPOSITION 2.3. *If* $\Sigma\in\mathbf{P}(\mathscr{K})$ *and* $A\in\mathbf{GL}(\mathscr{K})$, *then* $A\Sigma A^t\in\mathbf{P}(\mathscr{K})$.

PROOF. We will show that condition (ii) of Theorem 2.1 is valid with Σ replaced by $A\Sigma A^t$. Since $\Sigma\in\mathbf{P}(\mathscr{K})$, (ii) holds for Σ. Now replace x by $A^{-1}x$ in (ii) and let $B=A^{-1}$. The left-hand side of (ii) becomes $\operatorname{tr}((A\Sigma A^t)^{-1}xx^t)$ while the summands on the right-hand side become

$$\begin{aligned}&\operatorname{tr}\big(\Sigma^{-1}_{[K]\cdot}((Bx)_{[K]}-\Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}(Bx)_{\langle K\rangle})(\cdots)^t\big)\\&\quad=\operatorname{tr}\big(\Sigma^{-1}_{[K]\cdot}(B_{[K]}x_{[K]}+B_{[K\rangle}x_{\langle K\rangle}-\Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}B_{\langle K\rangle}x_{\langle K\rangle})(\cdots)^t\big)\\&\quad=\operatorname{tr}\big(B^t_{[K]}\Sigma^{-1}_{[K]\cdot}B_{[K]}\\&\qquad\times\big(x_{[K]}+(B^{-1}_{[K]}B_{[K\rangle}-B^{-1}_{[K]}\Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}B_{\langle K\rangle})x_{\langle K\rangle}\big)(\cdots)^t\big)\\&\quad=\operatorname{tr}\big((A_{[K]}\Sigma_{[K]\cdot}A^t_{[K]})^{-1}\\&\qquad\times\big(x_{[K]}-(A_{[K\rangle}A^{-1}_{\langle K\rangle}+A_{[K]}\Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}A^{-1}_{\langle K\rangle})x_{\langle K\rangle}\big)(\cdots)^t\\&\quad=\operatorname{tr}\big((A\Sigma A^t)^{-1}_{[K]\cdot}\big(x_{[K]}-(A\Sigma A^t)_{[K\rangle}(A\Sigma A^t)^{-1}_{\langle K\rangle}x_{\langle K\rangle}\big)(\cdots)^t\big).\end{aligned}$$

The first equality uses (2.18) and Proposition 2.2(ii), the third uses (2.22) and (2.23), and the fourth uses (2.27) and (2.28). Therefore, condition (ii) of Theorem 2.1 holds for $A\Sigma A^t$. □

2.5. *The $\mathscr{K}$-parametrization of* $\mathbf{P}(\mathscr{K})$. Theorem 2.2 establishes the one-to-one correspondence between Σ and its $\mathscr{K}$-parameters. Together with Theorem 2.1(ii) and Lemma 2.5, this decomposition of the parameter space $\mathbf{P}(\mathscr{K})$ yields the fundamental factorization of the likelihood function for the CI model $\mathbf{N}(\mathscr{K})$ (cf. Theorem 3.1).

LEMMA 2.4. *For any family*

$$\big((R_{[K\rangle}, \Lambda_{[K]})|K \in J(\mathscr{K})\big) \in \times(\mathbf{M}([K] \times \langle K\rangle) \times \mathbf{P}([K])|K \in J(\mathscr{K})),$$

there exists a matrix $A \in \mathbf{GL}(\mathscr{K})$ *such that for every* $K \in J(\mathscr{K})$,

$$A_{[K\rangle} = R_{[K\rangle}A_{\langle K\rangle}, \tag{2.29}$$

$$A_{[K]}A^t_{[K]} = \Lambda_{[K]}. \tag{2.30}$$

PROOF. First choose matrices $A_{[K]} \in \mathbf{GL}([K])$, $K \in J(\mathscr{K})$, that satisfy (2.30). As in Remark 2.1 let $K_1, \dots, K_q$ be a never-decreasing listing of the elements in $J(\mathscr{K})$. For notational convenience abbreviate K_k by k, $\langle K_k\rangle$ by $\langle k\rangle$, $[K_k\rangle$ by $[k\rangle$ and $[K_k]$ by $[k]$ whenever they appear as subscripts. If $K_1 \subset K_2$, then $\langle K_2\rangle = [K_1]$, so $A_{\langle 2\rangle} = A_{[1]}$ and $A_{[2\rangle}$ is uniquely determined by (2.29); if $K_1 \not\subseteq K_2$, then $\langle K_2\rangle = \varnothing$, so (2.29) is vacuous. Now suppose that we have determined $A_{[2\rangle}, \dots, A_{[k-1\rangle}$ satisfying (2.29). These $k-2$ matrices (some of which may be vacuous), together with $A_{[1]}, \dots, A_{[k-1]}$, completely determine $A_{\langle k\rangle}$. This follows from the decomposition [cf. (2.8)]

$$\langle K_k\rangle = \dot{\bigcup}([K_i]|K_i \subseteq \langle K_k\rangle) \tag{2.31}$$

and the fact that $K_i \subseteq \langle K_k\rangle \Rightarrow i < k$ for a never-decreasing listing. Now $A_{[k\rangle}$ is uniquely determined by (2.29) and, after induction on k, the matrix A is completely determined. By the surjectivity of the mapping (2.19), $A \in \mathbf{GL}(\mathscr{K})$. □

THEOREM 2.2 [The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$]. *The following mapping is bijective*:

$$\begin{aligned} \mathbf{P}(\mathscr{K}) &\to \times(\mathbf{M}([K] \times \langle K\rangle) \times \mathbf{P}([K])|K \in J(\mathscr{K})), \\ \Sigma &\to \big((\Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}, \Sigma_{[K]})|K \in J(\mathscr{K})\big). \end{aligned} \tag{2.32}$$

PROOF. By Theorem 2.1(ii), (2.32) is injective. To show that (2.32) is surjective, consider

$$\big((R_{[K\rangle}, \Lambda_{[K]})|K \in J(\mathscr{K})\big) \in \times(\mathbf{M}([K] \times \langle K\rangle) \times \mathbf{P}([K])|K \in J(\mathscr{K})).$$

By Lemma 2.4 there exists a matrix $A \in \mathbf{GL}(\mathscr{K})$ satisfying (2.29) and (2.30). Define $\Sigma := AA^t$; then $\Sigma \in \mathbf{P}(\mathscr{K})$ by Proposition 2.3 (with $\Sigma = 1_I$). The

$\mathscr{K}$-parameters of Σ are given by $\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1} = A_{[K\rangle}A_{\langle K\rangle}^{-1} = R_{[K\rangle}$ and $\Sigma_{[K]\cdot} = A_{[K]}A_{[K]}^t = \Lambda_{[K]}$, $K \in J(\mathscr{K})$ [set $\Sigma = 1_I$ in (2.27) and (2.28)]. □

2.6. *Transitive action of the group of $\mathscr{K}$-preserving matrices.*

THEOREM 2.3. *The action*

$$\begin{aligned} \mathbf{GL}(\mathscr{K}) \times \mathbf{P}(\mathscr{K}) &\to \mathbf{P}(\mathscr{K}), \\ (A, \Sigma) &\to A\Sigma A^t \end{aligned} \tag{2.33}$$

is well defined, transitive, continuous and proper.

PROOF. That (2.33) is well defined follows from Proposition 2.3. By Lemma 2.3 the bijective mapping (2.32) commutes with the actions (2.33) and (2.26). By Lemma 2.4, however, the action (2.26) is transitive, so it follows that (2.33) is also transitive. That (2.33) is continuous is trivial. Since $\mathbf{P}(\mathscr{K})$ and $\mathbf{GL}(\mathscr{K})$ are closed subsets of $\mathbf{P}(I)$ and $\mathbf{GL}(I)$, respectively, and the classical action of $\mathbf{GL}(I)$ on $\mathbf{P}(I)$ is proper, it follows that the action (2.33) is also proper. □

REMARK 2.3. Set $\mathbf{P}(\mathscr{K})^{-1} := \{\Delta^{-1} \in \mathbf{P}(I) | \Delta \in \mathbf{P}(\mathscr{K})\}$. By Theorem 2.3, the action

$$\begin{aligned} \mathbf{GL}(\mathscr{K}) \times \mathbf{P}(\mathscr{K})^{-1} &\to \mathbf{P}(\mathscr{K})^{-1}, \\ (A, \Delta) &\to (A^{-1})^t \Delta A^{-1} \end{aligned} \tag{2.34}$$

induced on $\mathbf{P}(\mathscr{K})^{-1}$ by (2.33) is also well defined, transitive, continuous and proper.

REMARK 2.4. Since both $\mathbf{P}(\mathscr{K})$ and $\mathbf{P}(\mathscr{K})^{-1}$ contain the $I \times I$ identity matrix 1_I, it follows from the transitivity of the actions (2.33) and (2.34) that

$$\mathbf{P}(\mathscr{K}) = \{AA^t \in \mathbf{P}(I) | A \in \mathbf{GL}(\mathscr{K})\}, \tag{2.35}$$

$$\mathbf{P}(\mathscr{K})^{-1} = \{A^tA \in \mathbf{P}(I) | A \in \mathbf{GL}(\mathscr{K})\}. \tag{2.36}$$

If $\mathscr{K} = \{\varnothing, I\}$, then $\mathbf{P}(\mathscr{K}) = \mathbf{P}(\mathscr{K})^{-1} = \mathbf{P}(I)$, so both actions (2.33) and (2.34) reduce to the well-known transitive actions of $\mathbf{GL}(I)$ on $\mathbf{P}(I)$. If $\mathscr{K}$ is a chain as in Examples 2.1 and 2.2, then again $\mathbf{P}(\mathscr{K}) = \mathbf{P}(\mathscr{K})^{-1} = \mathbf{P}(I)$, but now $\mathbf{GL}(\mathscr{K})$ is a group of nonsingular lower block-triangular matrices in the usual sense and the actions (2.33) and (2.34) are the well-known transitive actions of $\mathbf{GL}(\mathscr{K})$ on $\mathbf{P}(I)$.

The following lemma generalizes the Schur decomposition formula for $\det(\Sigma)$.

LEMMA 2.5. *For $\Sigma \in \mathbf{P}(\mathscr{K})$,*

$$\det(\Sigma) = \prod \big(\det(\Sigma_{[K]\cdot}) | K \in J(\mathscr{K})\big). \tag{2.37}$$

PROOF. By Theorem 2.3 there exists $A \in \mathbf{GL}(\mathscr{K})$ such that $\Sigma = AA^t$. Thus

$$\begin{aligned}\det(\Sigma) &= \det(AA^t)\\ &= \prod\left(\det\left(A_{[K]}A^t_{[K]}\right)|K \in J(\mathscr{K})\right)\\ &= \prod\left(\det(\Sigma_{[K]\cdot})|K \in J(\mathscr{K})\right).\end{aligned}$$

The second equality holds since A can be represented as a lower block-triangular matrix (cf. Remark 2.1), while the third equality follows from (2.28). □

2.7. *Reconstruction of* Σ *from its* $\mathscr{K}$*-parameters.* By Theorem 2.2, $\Sigma \in \mathbf{P}(\mathscr{K})$ is uniquely determined by its $\mathscr{K}$-parameters

$$\left((R_{[K\rangle}, \Lambda_{[K]})|K \in J(\mathscr{K})\right) \in \times(\mathbf{M}([K] \times \langle K\rangle) \times \mathbf{P}([K])|K \in J(\mathscr{K})),$$

where

$$R_{[K\rangle} = \Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle} \quad \text{and} \quad \Lambda_{[K]} = \Sigma_{[K]\cdot}. \tag{2.38}$$

Because the MLE $\hat{\Sigma}$ is obtained by first estimating the $\mathscr{K}$-parameters of Σ, then using these estimates to obtain $\hat{\Sigma}$ itself, it is important to find an explicit method for reconstructing $\Sigma \in \mathbf{P}(\mathscr{K})$ from its $\mathscr{K}$-parameters.

One such method is to apply the formula

$$\Sigma^{-1} = \sum(\Delta_I(K)|K \in J(\mathscr{K})), \tag{2.39}$$

which is just a reexpression of Theorem 2.1(ii), where $\Delta_I(K)$ is the $I \times I$ matrix whose $K \times K$ submatrix is

$$\begin{pmatrix} R^t_{[K\rangle}\Lambda^{-1}_{[K]}R_{[K\rangle} & -R^t_{[K\rangle}\Lambda^{-1}_{[K]} \\ -\Lambda^{-1}_{[K]}R_{[K\rangle} & \Lambda^{-1}_{[K]} \end{pmatrix} \tag{2.40}$$

and whose remaining entries are 0. In general, however, it is not a simple task to determine Σ from (2.39) by matrix inversion. We now present a stepwise algorithm for reconstructing Σ directly from its $\mathscr{K}$-parameters.

Let $K_1, \dots, K_q$ be a never-decreasing listing of the members of the poset $J(\mathscr{K})$ (cf. Remark 2.1 and the proof of Lemma 2.4), partition Σ according to (2.9) and list the $\mathscr{K}$-parameters in the corresponding order:

$$\begin{aligned}&(\Lambda_{[1]}, R_{[2\rangle}, \Lambda_{[2]}, \dots, R_{[q\rangle}, \Lambda_{[q]})\\ &\quad\in \mathbf{P}([K_1]) \times \mathbf{M}([K_2] \times \langle K_2\rangle) \times \mathbf{P}([K_2])\\ &\qquad\times \cdots \times \mathbf{M}\left([K_q] \times \langle K_q\rangle\right) \times \mathbf{P}\left([K_q]\right).\end{aligned} \tag{2.41}$$

(Recall that whenever they appear as subscripts, K_k, $\langle K_k\rangle$, $[K_k\rangle$ and $[K_k]$ are abbreviated by k, $\langle k\rangle$, $[k\rangle$ and $[k]$, respectively.) The reconstruction algorithm proceeds stepwise as follows. At step k the relations in (2.38) are inverted to determine $\Sigma_{[k\rangle}$ and $\Sigma_{[k]\cdot}$ from the corresponding $\mathscr{K}$-parameters $R_{[k\rangle}$ and $\Lambda_{[k]}$ and from the matrix $\Sigma_{1\cup\cdots\cup(k-1)}$ constructed in step $k-1$. The remaining entries in $\Sigma_{1\cup\cdots\cup k}$ are determined by the CI conditions.

Step 1: $$\Sigma_{[1]} = \Lambda_{[1]};$$

Step 2: $$\Sigma_{[2\rangle} = R_{[2\rangle}\Sigma_{\langle 2\rangle},$$
$$\Sigma_{[2]} = \Lambda_{[2]} + R_{[2\rangle}\Sigma_{\langle 2]}.$$

At this point the submatrix $\Sigma_{1\cup 2}$ is completely determined: If $K_1 \subset K_2$, then $\Sigma_{1\cup 2} = \Sigma_2$, whereas if $K_1 \not\subseteq K_2$, then $K_1 \cap K_2 = \varnothing$, so the $[K_1] \times [K_2]$-submatrix of Σ is 0 by (2.2). (Recall that $1 \cup 2$ abbreviates $K_1 \cup K_2$ when appearing as a subscript.) By (2.42), $\langle K_3\rangle \subseteq K_1 \cup K_2$, so $\Sigma_{\langle 3\rangle}$ is a submatrix of $\Sigma_{1\cup 2}$; hence the next step may be carried out.

Step 3a: $$\Sigma_{[3\rangle} = R_{[3\rangle}\Sigma_{\langle 3\rangle},$$
$$\Sigma_{[3]} = \Lambda_{[3]} + R_{[3\rangle}\Sigma_{\langle 3]}.$$

It is important to note that after steps 1, 2 and 3a, the three submatrices Σ_1, Σ_2 and Σ_3 are now determined but the complete submatrix $\Sigma_{1\cup 2\cup 3}$ may not yet be fully determined. The remaining $[K_3] \times ((K_1 \cup K_2 \cup K_3)\setminus K_3)$-submatrix of $\Sigma_{1\cup 2\cup 3}$, which we denote by $\Sigma_{[3\}}$, is determined from $\Sigma_{1\cup 2}$ by means of the pairwise CI requirements imposed by $\mathscr{K}$ [cf. (2.44)]:

Step 3b: $$\Sigma_{[3\}} = R_{[3\rangle}\Sigma_{\langle 3\}}$$
$$\left(= \Sigma_{[3\rangle}\Sigma_{\langle 3\rangle}^{-1}\Sigma_{\langle 3\}}\right),$$

where $\Sigma_{\langle 3\}}$ is the $\langle K_3\rangle \times ((K_1 \cup K_2 \cup K_3)\setminus K_3)$-submatrix of $\Sigma_{1\cup 2\cup 3}$. By (2.42) and (2.43), however, $\Sigma_{\langle 3\}}$ is in fact a submatrix of $\Sigma_{1\cup 2}$; hence it may be used to obtain $\Sigma_{[3\}}$ in this step.

After $k-1$ such steps, the submatrix $\Sigma_{1\cup\cdots\cup(k-1)}$ is fully determined and in turn may be used to obtain $\Sigma_{1\cup\cdots\cup k}$ as follows. First note that by Proposition 2.1 and the never-decreasing nature of $K_1,\dots,K_q$,

$$K_1 \cup \cdots \cup K_k = \dot{\bigcup}\left([K_j] | j = 1,\dots,k\right),$$
$$K_k = \dot{\bigcup}\left([K_j] | j = 1,\dots,k,\ K_j \subseteq K_k\right).$$

From these relations and (2.3) it may be deduced that

$$\langle K_k\rangle = K_k \cap (K_1 \cup \cdots \cup K_{k-1}) \subseteq K_1 \cup \cdots \cup K_{k-1}, \tag{2.42}$$

$$\begin{aligned}(K_1 \cup \cdots \cup K_k)\setminus K_k &= (K_1 \cup \cdots \cup K_{k-1})\setminus\langle K_k\rangle\\ &\subseteq K_1 \cup \cdots \cup K_{k-1}.\end{aligned} \tag{2.43}$$

Thus, if we denote the $[K_k] \times ((K_1 \cup \cdots \cup K_k)\setminus K_k)$-submatrix of $\Sigma_{1\cup\cdots\cup k}$ by $\Sigma_{[k\}}$ and the $\langle K_k\rangle \times ((K_1 \cup \cdots \cup K_k)\setminus K_k)$-submatrix by $\Sigma_{\langle k\}}$, it follows from (2.42) and (2.43) that both $\Sigma_{\langle k\rangle}$ and $\Sigma_{\langle k\}}$ are in fact submatrices of $\Sigma_{1\cup\cdots\cup(k-1)}$, so the next step may be carried out:

Step k:
$$\begin{aligned}\Sigma_{[k\rangle} &= R_{[k\rangle}\Sigma_{\langle k\rangle},\\ \Sigma_{[k]} &= \Lambda_{[k]} + R_{[k\rangle}\Sigma_{\langle k]},\\ \Sigma_{[k\}} &= R_{[k\}}\Sigma_{\langle k\}}\\ &\left(= \Sigma_{[k\}}\Sigma_{\langle k\rangle}^{-1}\Sigma_{\langle k\}}\right).\end{aligned} \tag{2.44}$$

The relation in (2.44) is seen as follows. Since K_k (=: L) and $K_1 \cup \cdots \cup K_{k-1}$ (=: M) are members of $\mathscr{K}$, it follows from the pairwise CI condition (2.2) and from (2.42) and (2.43) that the $[K_k] \times ((K_1 \cup \cdots \cup K_k) \setminus K_k)$-submatrix of $(\Sigma_{1\cup \cdots \cup k})^{-1}$ is a zero matrix, which is equivalent to (2.44).

The submatrix $\Sigma_{1\cup \cdots \cup k}$ of Σ is fully determined after step k; after q steps, $\Sigma_{1\cup \cdots \cup q} \equiv \Sigma$ is fully determined.

[In carrying out this algorithm, one must use the convention that if $C \neq \varnothing$ and $D \neq \varnothing$, then the product of a $C \times \varnothing$ matrix with a $\varnothing \times D$ matrix is the $C \times D$ zero matrix.]

2.8. *Examples.* A series of nine examples will illustrate the following basic aspects of a lattice CI model $\mathbf{N}(\mathscr{K})$: (a) the distributive lattice $\mathscr{K} \subseteq \mathscr{D}(I)$ and the poset $J(\mathscr{K})$ of join-irreducible elements; (b) the $\mathscr{K}$-parametrization (2.32) of $\mathbf{P}(\mathscr{K})$ and the associated decomposition of $\operatorname{tr}(\Sigma^{-1}xx^t)$ given in Theorem 2.1(ii); (c) the choice of a never-decreasing listing of the members of $J(\mathscr{K})$ and the reconstruction of the covariance matrix $\Sigma \in \mathbf{P}(\mathscr{K})$ from its ordered $\mathscr{K}$-parameters [cf. (2.38)] by means of the stepwise algorithm in Section 2.7, as well as the form of the precision matrix $\Delta \equiv \Sigma^{-1} \in \mathbf{P}(\mathscr{K})^{-1}$; (d) the form of the $\mathscr{K}$-preserving matrices, that is, the group $\mathbf{GL}(\mathscr{K})$ of matrices, partitioned according the ordered decomposition (2.15), that acts transitively on $\mathbf{P}(\mathscr{K})$ (cf. Remarks 2.1 and 2.2). The reader should verify directly that (2.35) and (2.36) hold for $\mathbf{P}(\mathscr{K})$ and $\mathbf{GL}(\mathscr{K})$ in these nine examples.

In each example the lattice diagram of $\mathscr{K}$ appears in an accompanying figure, in which the members of $J(\mathscr{K})$ are indicated by open circles and the remaining members of $\mathscr{K}$ by solid dots. In each figure the minimal element $\varnothing$ appears at the left while the maximal element I appears at the right.

These examples will be continued in Section 3.2, where the MLE $\hat{\Sigma}$ is determined for each of these CI models, and in Andersson and Perlman (1993b) to provide examples of the problem of testing one CI model against another. Additional examples appear in Andersson and Perlman (1991).

EXAMPLE 2.1. First consider the simple case where $\mathscr{K} = \{\varnothing, L, I\}$ (see Figure 1). Since $\mathscr{K}$ is a chain, $\mathbf{P}(\mathscr{K}) = \mathbf{P}(I)$. Note that $J(\mathscr{K}) = \{L, I\}$ and $\langle L \rangle = \varnothing$, $\langle I \rangle = [L] = L$. Thus the $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ becomes

$$(2.45)\qquad \begin{aligned} \mathbf{P}(I) &\leftrightarrow \mathbf{P}(L) \times \mathbf{M}([I] \times L) \times \mathbf{P}([I]), \\ \Sigma &\leftrightarrow \left(\Sigma_L, \Sigma_{[I\rangle}\Sigma_L^{-1}, \Sigma_{[I]\cdot}\right), \end{aligned}$$

and

$$(2.46)\qquad \operatorname{tr}(\Sigma^{-1}xx^t) = \operatorname{tr}\left(\Sigma_L^{-1}x_L x_L^t\right) + \operatorname{tr}\left(\Sigma_{[I]\cdot}^{-1}\left(x_{[I]} - \Sigma_{[I\rangle}\Sigma_L^{-1}x_L\right)(\cdots)^t\right).$$

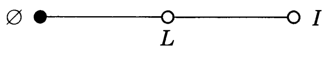

FIG. 1.

FIG. 2.

The algorithm for reconstructing Σ from its ordered $\mathscr{K}$-parameters $\Lambda_{[L]}$, $R_{[I\rangle}$ and $\Lambda_{[I]}$ takes the following form:

Step 1: $\Sigma_L = \Lambda_{[L]}$;

Step 2: $\Sigma_{[I\rangle} = R_{[I\rangle}\Sigma_L$,

$\Sigma_{[I]} = \Lambda_{[I]} + R_{[I\rangle}\Sigma_{\langle I]}$.

The group $\mathbf{GL}(\mathscr{K})$ is a lower block-triangular matrix group in the ordinary sense: $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$A = \begin{pmatrix} A_L & 0 \\ A_{[L\rangle} & A_{[I]} \end{pmatrix}. \tag{2.47}$$

EXAMPLE 2.2. If $\mathscr{K} \equiv \{\varnothing \equiv K_0, K_1, \dots, K_{q-1}, K_q \equiv I\}$ is an ascending chain, that is, $\varnothing \subset K_1 \subset \cdots \subset K_{q-1} \subset I$, then a well-known generalization of the preceding example is obtained (see Figure 2). Again $\mathbf{P}(\mathscr{K}) = \mathbf{P}(I)$, but the $\mathscr{K}$-parametrization is changed. Note that $J(\mathscr{K}) = \{K_1, \dots, K_q\}$ and $\langle K_1 \rangle = \varnothing$, $\langle K_k \rangle = K_{k-1}$, $k = 2, \dots, q$. Then the $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ becomes

$$\begin{aligned} \mathbf{P}(I) &\leftrightarrow \mathbf{P}(K_1) \times \mathbf{M}([K_2] \times K_1) \times \mathbf{P}([K_2]) \\ &\qquad \times \cdots \times \mathbf{M}\big([K_q] \times K_{q-1}\big) \times \mathbf{P}\big([K_q]\big), \\ \Sigma &\leftrightarrow \big(\Sigma_1, \Sigma_{[2\rangle}\Sigma_1^{-1}, \Sigma_{[2]\cdot}, \dots, \Sigma_{[q\rangle}\Sigma_{q-1}^{-1}, \Sigma_{[q]\cdot}\big), \end{aligned} \tag{2.48}$$

and

$$\begin{aligned} \operatorname{tr}(\Sigma^{-1}xx^t) &= \operatorname{tr}\big(\Sigma_1^{-1}x_1x_1^t\big) + \operatorname{tr}\Big(\Sigma_{[2]\cdot}^{-1}\big(x_{[2]} - \Sigma_{[2\rangle}\Sigma_1^{-1}x_1\big)(\cdots)^t\Big) \\ &\quad + \cdots + \operatorname{tr}\Big(\Sigma_{[q]\cdot}^{-1}\big(x_{[q]} - \Sigma_{[q\rangle}\Sigma_{q-1}^{-1}x_{q-1}\big)(\cdots)^t\Big), \end{aligned} \tag{2.49}$$

where $K_1, K_2, \dots, K_q$ are abbreviated as $1, 2, \dots, q$ whenever they occur as subscripts. Then Σ is reconstructed from its ordered $\mathscr{K}$-parameters $\Lambda_{[1]}, R_{[2\rangle}, \Lambda_{[2]}, \dots, R_{[q\rangle}, \Lambda_{[q]}$ as follows:

Step 1: $\Sigma_1 = \Lambda_{[1]}$;

Step 2: $\Sigma_{[2\rangle} = R_{[2\rangle}\Sigma_1$,

$\Sigma_{[2]} = \Lambda_{[2]} + R_{[2\rangle}\Sigma_{\langle 2]}$;

$\vdots$

Step q: $\Sigma_{[q\rangle} = R_{[q\rangle}\Sigma_{q-1}$,

$\Sigma_{[q]} = \Lambda_{[q]} + R_{[q\rangle}\Sigma_{\langle q]}$.

The group $\mathbf{GL}(\mathscr{K})$ is again a group of lower block-triangular matrices in the usual sense. For example, when $q = 4$, $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$A = \left(\begin{array}{cc:c:c} A_1 & 0 & 0 & 0 \\ A_{[2\rangle} & A_{[2]} & 0 & 0 \\ \hdashline \multicolumn{2}{c:}{A_{[3\rangle}} & A_{[3]} & 0 \\ \hdashline \multicolumn{3}{c:}{A_{[4\rangle}} & A_{[4]} \end{array}\right). \tag{2.50}$$

EXAMPLE 2.3. Consider the lattice $\mathscr{K} = \{\varnothing \equiv L \cap M, L, M, L \cap M \equiv I\}$ (see Figure 3). Here the CI requirement determined by $\mathscr{K}$ is nontrivial, so $\mathbf{P}(\mathscr{K}) \subset \mathbf{P}(I)$. Now $J(\mathscr{K}) = \{L, M\}$ and $\langle L \rangle = \langle M \rangle = \varnothing$. The $\mathscr{K}$-parametrization takes the form

$$\begin{gathered} \mathbf{P}(\mathscr{K}) \leftrightarrow \mathbf{P}(L) \times \mathbf{P}(M), \\ \Sigma \leftrightarrow (\Sigma_L, \Sigma_M) \end{gathered} \tag{2.51}$$

and

$$\operatorname{tr}(\Sigma^{-1}xx^t) = \operatorname{tr}\left(\Sigma_L^{-1}x_L x_L^t\right) + \operatorname{tr}\left(\Sigma_M^{-1}x_M x_M^t\right). \tag{2.52}$$

Since L, M is a never-decreasing listing of $J(\mathscr{K})$, Σ may be reconstructed from its ordered nontrivial $\mathscr{K}$-parameters $\Lambda_{[L]}, \Lambda_{[M]}$ as follows:

Step 1: $\Sigma_L = \Lambda_{[L]}$;

Step 2: $\Sigma_M = \Lambda_{[M]}$,

$\Sigma_{[M\rangle} = 0.$

Thus $\mathbf{P}(\mathscr{K})$ consists of all block-diagonal matrices Σ of the form

$$\Sigma = \begin{pmatrix} \Sigma_L & 0 \\ 0 & \Sigma_M \end{pmatrix}, \tag{2.53}$$

where Σ is partitioned according to the ordered decomposition

$$I = L \mathbin{\dot{\cup}} M. \tag{2.54}$$

In this example, as in Examples 2.1 and 2.2, $\mathbf{P}(\mathscr{K}) = \mathbf{P}(\mathscr{K})^{-1}$ and both are *linear*, that is, closed under (nonnegative) linear combinations. The group

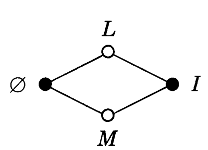

FIG. 3.

$\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$A = \begin{pmatrix} A_L & 0 \\ 0 & A_M \end{pmatrix}. \tag{2.55}$$

EXAMPLE 2.4. If $\mathscr{K} = \{\varnothing \equiv L \cap M, L, M, L \cup M, I\}$ (see Figure 4), then again $\mathbf{P}(\mathscr{K}) \subset \mathbf{P}(I)$. Here $J(\mathscr{K}) = \{L, M, I\}$ and $\langle L \rangle = \langle M \rangle = \varnothing$, $\langle I \rangle = L \cup M$. The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ assumes the form

$$\begin{aligned} \mathbf{P}(\mathscr{K}) &\leftrightarrow \mathbf{P}(L) \times \mathbf{P}(M) \times \mathbf{M}([I] \times (L \cup M)) \times \mathbf{P}([I]), \\ \Sigma &\leftrightarrow \left(\Sigma_L, \Sigma_M, \Sigma_{[I\rangle}\Sigma_{L \cup M}^{-1}, \Sigma_{[I]\cdot}\right) \end{aligned} \tag{2.56}$$

and

$$\begin{aligned} &\operatorname{tr}(\Sigma^{-1}xx^t) \\ &\quad = \operatorname{tr}\left(\Sigma_L^{-1}x_L x_L^t\right) + \operatorname{tr}\left(\Sigma_M^{-1}x_M x_M^t\right) \\ &\qquad + \operatorname{tr}\left(\Sigma_{[I]\cdot}^{-1}\left(x_{[I]} - \Sigma_{[I\rangle}\Sigma_{L \cup M}^{-1}x_{L \cup M}\right)(\cdots)^t\right). \end{aligned}$$

Now L, M, I is a never-decreasing listing of $J(\mathscr{K})$, so Σ may be reconstructed from its ordered nontrivial $\mathscr{K}$-parameters $\Lambda_{[L]}, \Lambda_{[M]}, R_{[I\rangle}, \Lambda_{[I]}$ as follows:

Steps 1 *and* 2: Repeat steps 1 and 2 in Example 2.3.

Step 3:

$$\begin{aligned} \Sigma_{[I\rangle} &= R_{[I\rangle}\operatorname{Diag}(\Sigma_L, \Sigma_M), \\ \Sigma_{[I]} &= \Lambda_{[I]} + R_{[I\rangle}\Sigma_{\langle I]}. \end{aligned}$$

Thus $\mathbf{P}(\mathscr{K})$ consists of all Σ of the form

$$\Sigma = \left(\begin{array}{cc:c} \Sigma_L & 0 & \\ 0 & \Sigma_M & \Sigma_{\langle I]} \\ \hdashline \multicolumn{2}{c:}{\Sigma_{[I\rangle}} & \Sigma_{[I]} \end{array}\right), \tag{2.57}$$

where Σ is partitioned according to the ordered decomposition

$$I = L \,\dot{\cup}\, M \,\dot{\cup}\, [I]. \tag{2.58}$$

The precision matrix $\Delta \equiv \Sigma^{-1} \in \mathbf{P}(\mathscr{K})^{-1}$ is characterized by the condition that $\Sigma_{L \cup M}^{-1} = \operatorname{Diag}(\Sigma_L^{-1}, \Sigma_M^{-1})$. Thus, unlike the preceding example, here $\mathbf{P}(\mathscr{K})$ is linear while $\mathbf{P}(\mathscr{K})^{-1}$ is not. The group $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular

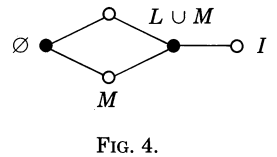

FIG. 4.

$I \times I$ matrices of the form

$$A = \begin{pmatrix} A_L & 0 & \vdots & 0 \\ 0 & A_M & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \multicolumn{2}{c}{A_{[I\rangle}} & \vdots & A_{[I]} \end{pmatrix}. \tag{2.59}$$

EXAMPLE 2.5. Suppose that $\mathscr{K} = \{\varnothing, L \cap M, L, M, L \cup M \equiv I\}$ (see Figure 5). [Note that (1.3) is a special case.] Now $J(\mathscr{K}) = \{L \cap M, L, M\}$, and $\langle L \cap M \rangle = \varnothing$, $\langle L \rangle = \langle M \rangle = L \cap M$. The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ is given by

$$\begin{aligned} \mathbf{P}(\mathscr{K}) &\leftrightarrow \mathbf{P}(L \cap M) \times \mathbf{M}([L] \times (L \cap M)) \\ &\qquad \times \mathbf{P}([L]) \times \mathbf{M}([M] \times (L \cap M)) \times \mathbf{P}([M]), \\ \Sigma &\leftrightarrow \left(\Sigma_{L \cap M}, \Sigma_{[L\rangle}\Sigma_{L \cap M}^{-1}, \Sigma_{[L]\cdot}, \Sigma_{[M\rangle}\Sigma_{L \cap M}^{-1}, \Sigma_{[M]\cdot}\right) \end{aligned} \tag{2.60}$$

and

$$\begin{aligned} \operatorname{tr}(\Sigma^{-1}xx^t) &= \operatorname{tr}\left(\Sigma_{L \cap M}^{-1} x_{L \cap M} x_{L \cap M}^t\right) \\ &\quad + \operatorname{tr}\left(\Sigma_{[L]\cdot}^{-1}\left(x_{[L]} - \Sigma_{[L\rangle}\Sigma_{L \cap M}^{-1} x_{L \cap M}\right)(\cdots)^t\right) \\ &\quad + \operatorname{tr}\left(\Sigma_{[M]\cdot}^{-1}\left(x_{[M]} - \Sigma_{[M\rangle}\Sigma_{L \cap M}^{-1} x_{L \cap M}\right)(\cdots)^t\right). \end{aligned} \tag{2.61}$$

Since $L \cap M, L, M$ is a never-decreasing listing of $J(\mathscr{K})$, Σ may be reconstructed from its ordered $\mathscr{K}$-parameters $\Lambda_{[L \cap M]}$, $R_{[L\rangle}$, $\Lambda_{[L]}$, $R_{[M\rangle}$, $\Lambda_{[M]}$ as follows:

Step 1: $\Sigma_{L \cap M} = \Lambda_{[L \cap M]}$;

Step 2: $\Sigma_{[L\rangle} = R_{[L\rangle}\Sigma_{L \cap M}$,

$\Sigma_{[L]} = \Lambda_{[L]} + R_{[L\rangle}\Sigma_{\langle L]}$,

Step 3: $\Sigma_{[M\rangle} = R_{[M\rangle}\Sigma_{L \cap M}$,

(2.62) $\Sigma_{[M]} = \Lambda_{[M]} + R_{[M\rangle}\Sigma_{\langle M]}$,

$\Sigma_{[M\not\rangle} = R_{[M\rangle}\Sigma_{\langle L]}$

$\left(= \Sigma_{[M\rangle}\Sigma_{L \cap M}^{-1}\Sigma_{\langle L]}\right)$.

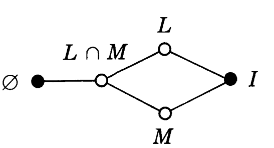

FIG. 5.

(Note that $\Sigma_{\langle M\}} = \Sigma_{\langle L]\cdot}$.) Thus $\mathbf{P}(\mathscr{K})$ consists of all $\Sigma \in \mathbf{P}(I)$ of the form

$$\Sigma = \begin{pmatrix} \Sigma_{L\cap M} & \Sigma_{\langle L]} & \Sigma_{\langle M]} \\ \Sigma_{[L\rangle} & \Sigma_{[L]} & \Sigma_{\{M]\cdot} \\ \Sigma_{[M\rangle} & \Sigma_{[M\}} & \Sigma_{[M]} \end{pmatrix}, \tag{2.63}$$

such that $\Sigma_{[M\}}$ satisfies (2.62) and where Σ is partitioned according to the ordered decomposition

$$I = (L \cap M) \dot{\cup} [L] \dot{\cup} [M]. \tag{2.64}$$

Then $\mathbf{P}(\mathscr{K})^{-1}$ consists of all $\Delta \in \mathbf{P}(I)$ having the simple form

$$\Delta = \begin{pmatrix} \Delta_{L\cap M} & \Delta_{\langle L]} & \Delta_{\langle M]} \\ \Delta_{[L\rangle} & \Delta_{[L]} & 0 \\ \Delta_{[M\rangle} & 0 & \Delta_{[M]} \end{pmatrix}. \tag{2.65}$$

Thus in this example $\mathbf{P}(\mathscr{K})^{-1}$ is linear while $\mathbf{P}(\mathscr{K})$ is not. The group $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$A = \begin{pmatrix} A_{L\cap M} & 0 & 0 \\ A_{[L\rangle} & A_{[L]} & 0 \\ A_{[M\rangle} & 0 & A_{[M]} \end{pmatrix}. \tag{2.66}$$

EXAMPLE 2.6. Consider the lattice $\mathscr{K} = \{\varnothing, L \cap M, L, M, L \cup M, I\}$ (see Figure 6). Note that $J(\mathscr{K}) = \{\varnothing, L \cap M, L, M, I\}$ and $\langle L \cap M\rangle = \varnothing$, $\langle L\rangle = \langle M\rangle = L \cap M$, $\langle I\rangle = L \cup M$. The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ is given by

$$\begin{aligned} \mathbf{P}(\mathscr{K}) \leftrightarrow\ & \mathbf{P}(L \cap M) \times \mathbf{M}([L] \times (L \cap M)) \times \mathbf{P}([L]) \\ & \times \mathbf{M}([M] \times (L \cap M)) \times \mathbf{P}([M]) \\ & \times \mathbf{M}([I] \times (L \cup M)) \times \mathbf{P}([I]), \\ \Sigma \leftrightarrow\ & \left(\Sigma_{L\cap M}, \Sigma_{[L\rangle}\Sigma_{L\cap M}^{-1}, \Sigma_{[L]\cdot}, \Sigma_{[M\rangle}\Sigma_{L\cap M}^{-1}, \Sigma_{[M]\cdot}, \Sigma_{[I\rangle}\Sigma_{L\cup M}^{-1}, \Sigma_{[I]\cdot}\right), \end{aligned} \tag{2.67}$$

and

$$\begin{aligned} \operatorname{tr}(\Sigma^{-1}xx^t) =\ & \operatorname{tr}\left(\Sigma_{L\cap M}^{-1} x_{L\cap M} x_{L\cap M}^t\right) \\ & + \operatorname{tr}\left(\Sigma_{[L]\cdot}^{-1}\left(x_{[L]} - \Sigma_{[L\rangle}\Sigma_{L\cap M}^{-1} x_{L\cap M}\right)(\cdots)^t\right) \\ & + \operatorname{tr}\left(\Sigma_{[M]\cdot}^{-1}\left(x_{[M]} - \Sigma_{[M\rangle}\Sigma_{L\cap M}^{-1} x_{L\cap M}\right)(\cdots)^t\right) \\ & + \operatorname{tr}\left(\Sigma_{[I]\cdot}^{-1}\left(x_{[I]} - \Sigma_{[I\rangle}\Sigma_{L\cup M}^{-1} x_{L\cup M}\right)(\cdots)^t\right). \end{aligned} \tag{2.68}$$

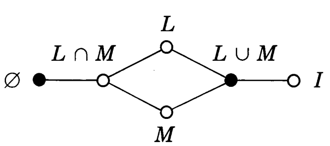

FIG. 6.

Since $L \cap M, L, M, I$ is a never-decreasing listing of $J(\mathscr{K})$, Σ can be reconstructed from $\Lambda_{[L\cap M]}, R_{[L\rangle}, \Lambda_{[L]}, R_{[M\rangle}, \Lambda_{[M]}, R_{[I\rangle}, \Lambda_{[I]}$ as follows:

Steps 1, 2, 3: Repeat steps 1, 2 and 3 in Example 2.5 to obtain $\Sigma_{L\cup M}$.

Step 4:

$$\Sigma_{[I\rangle} = R_{[I\rangle}\Sigma_{L\cup M}$$

$$\Sigma_{[I]} = \Lambda_{[I]} + R_{[I\rangle}\Sigma_{\langle I]}.$$

Thus $\mathbf{P}(\mathscr{K})$ consists of all Σ of the form

$$\Sigma = \begin{pmatrix} \Sigma_{L\cup M} & \Sigma_{\langle I]} \\ \Sigma_{[I\rangle} & \Sigma_{[I]} \end{pmatrix} \tag{2.69}$$

partitioned according to $I = I_{L\cup M}\,\dot{\cup}\,[I]$, where $\Sigma_{L\cup M}$ is given by (2.63), (2.64) and (2.62). The precision matrix $\Delta \equiv \Sigma^{-1} \in \mathbf{P}(\mathscr{K})^{-1}$ is characterized by the condition that $\Sigma_{L\cup M}^{-1}$ have the form (2.65). Thus *neither* $\mathbf{P}(\mathscr{K})$ *nor* $\mathbf{P}(\mathscr{K})^{-1}$ *is linear*. The group $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$A = \left(\begin{array}{ccc:c} A_{L\cap M} & 0 & 0 & 0 \\ A_{[L\rangle} & A_{[L]} & 0 & 0 \\ A_{[M\rangle} & 0 & A_{[M]} & 0 \\ \hdashline & A_{[I\rangle} & & A_{[I]} \end{array}\right). \tag{2.70}$$

EXAMPLE 2.7. Let $\mathscr{K}$ be the lattice in Figure 7. Then $J(\mathscr{K}) = \{L\cap M, L, M, L', M'\}$ and $\langle L\cap M\rangle = \varnothing$, $\langle L\rangle = \langle M\rangle = L\cap M$, $\langle L'\rangle = \langle M'\rangle = L\cup M \equiv L'\cap M'$. The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ is given by

$$\begin{aligned} \mathbf{P}(\mathscr{K}) \leftrightarrow\ & \mathbf{P}(L\cap M) \times \mathbf{M}([L]\times(L\cap M)) \times \mathbf{P}([L]) \\ & \times \mathbf{M}([M]\times(L\cap M)) \times \mathbf{P}([M]) \\ & \times \mathbf{M}([L']\times(L\cup M)) \times \mathbf{P}([L']) \\ & \times \mathbf{M}([M']\times(L\cup M)) \times \mathbf{P}([M']), \\ \Sigma \leftrightarrow\ & \big(\Sigma_{L\cap M}, \Sigma_{[L\rangle}\Sigma_{L\cap M}^{-1}, \Sigma_{[L]\cdot}, \Sigma_{[M\rangle}\Sigma_{L\cap M}^{-1}, \Sigma_{[M]\cdot}, \\ & \quad \Sigma_{[L'\rangle}\Sigma_{L\cup M}^{-1}, \Sigma_{[L']\cdot}, \Sigma_{[M'\rangle}\Sigma_{L\cup M}^{-1}, \Sigma_{[M']\cdot}\big), \end{aligned} \tag{2.71}$$

from which the decomposition of $\mathrm{tr}(\Sigma^{-1}xx^t)$ is directly obtained. The matrix Σ

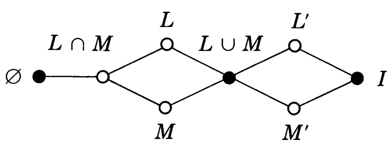

FIG. 7.

can be reconstructed from its ordered $\mathscr{K}$-parameters

$$\Lambda_{[L\cap M]}, R_{[L\rangle}, \Lambda_{[L]}, R_{[M\rangle}, \Lambda_{[M]}, R_{[L'\rangle}, \Lambda_{[L'\rangle}, R_{[M'\rangle}, \Lambda_{[M']}$$

as follows:

Steps 1, 2, *and* 3: Repeat steps 1, 2 and 3 in Example 2.5 to obtain $\Sigma_{L\cup M} \equiv \Sigma_{L'\cap M'}$.

Steps 4 *and* 5: Repeat steps 2 and 3 in Example 2.5 with L, M replaced by L', M'.

Thus $\mathbf{P}(\mathscr{K})$ consists of all Σ of the form (2.63) with L, M replaced by L', M', partitioned according to the ordered decomposition

$$I = (L \cup M)\, \dot{\cup} [L']\, \dot{\cup} [M'] \tag{2.72}$$

and where $\Sigma_{L'\cap M'} \equiv \Sigma_{L\cup M}$ is given by (2.63). The precision matrix $\Delta \equiv \Sigma^{-1}$ has the form (2.65) with L, M replaced by L', M' and satisfies the condition that $\Sigma_{L\cup M}^{-1}$ has the form (2.65). Again, neither $\mathbf{P}(\mathscr{K})$ nor $\mathbf{P}(\mathscr{K})^{-1}$ is linear. The group $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$A = \left(\begin{array}{ccc:cc} A_{L\cap M} & 0 & 0 & 0 & 0 \\ A_{[L\rangle} & A_{[L]} & 0 & 0 & 0 \\ A_{[M\rangle} & 0 & A_{[M]} & 0 & 0 \\ \hdashline & A_{[L'\rangle} & & A_{[L']} & 0 \\ & A_{[M'\rangle} & & 0 & A_{[M']} \end{array}\right). \tag{2.73}$$

EXAMPLE 2.8. Let $\mathscr{K}$ be the lattice in Figure 8. Here $J(\mathscr{K}) = \{L\cap M, L, M, L'', M'\}$ and $\langle L\cap M\rangle = \varnothing$, $\langle L\rangle = \langle M\rangle = L\cap M$, $\langle L''\rangle = L$, $\langle M'\rangle = L\cup M = L'\cap M'$. The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ is given by

$$\begin{aligned} \mathbf{P}(I) \leftrightarrow\ & \mathbf{P}(L\cap M) \times \mathbf{M}([L]\times(L\cap M)) \times \mathbf{P}([L]) \\ & \times \mathbf{M}([M]\times(L\cap M)) \times \mathbf{P}([M]) \\ & \times \mathbf{M}([L'']\times L) \times \mathbf{P}([L'']) \\ & \times \mathbf{M}([M']\times(L\cup M)) \times \mathbf{P}([M']), \\ \Sigma \leftrightarrow\ & \big(\Sigma_{L\cap M}, \Sigma_{[L\rangle}\Sigma_{L\cap M}^{-1}, \Sigma_{[L]\cdot}, \Sigma_{[M\rangle}\Sigma_{L\cap M}^{-1}, \Sigma_{[M]\cdot}, \\ & \Sigma_{[L''\rangle}\Sigma_L^{-1}, \Sigma_{[L'']\cdot}, \Sigma_{[M'\rangle}\Sigma_{L\cup M}^{-1}, \Sigma_{[M']\cdot}\big), \end{aligned} \tag{2.74}$$

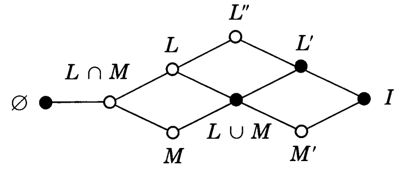

FIG. 8.

from which the decomposition of $\operatorname{tr}(\Sigma^{-1}xx^t)$ follows directly. The matrix Σ can be reconstructed from its ordered $\mathscr{K}$-parameters

$$\Lambda_{[L\cap M]},\ R_{[L\rangle},\ \Lambda_{[L]},\ R_{[M\rangle},\ \Lambda_{[M]},\ R_{[L''\rangle},\ \Lambda_{[L'']},\ R_{[M'\rangle},\ \Lambda_{[M']}$$

as follows:

Steps 1, 2, 3: Repeat steps 1, 2 and 3 in Example 2.5, to obtain $\Sigma_{L\cup M}\equiv\Sigma_{L'\cap M'}$.

Step 4:

$$\begin{aligned}
&\Sigma_{[L''\rangle}=R_{[L''\rangle}\Sigma_L,\\
&\Sigma_{[L'']}=\Lambda_{[L'']}+R_{[L''\rangle}\Sigma_{\langle L'']},\\
(2.75)\qquad&\Sigma_{[L''\}}=R_{[L''\rangle}\Sigma_{\langle L''\}}\\
&\left(=\Sigma_{[L''\rangle}\Sigma_L^{-1}\begin{pmatrix}\Sigma_{\langle M]}\\ \Sigma_{\{M]}\end{pmatrix}\right),
\end{aligned}$$

where $\Sigma_{\{M]}=\Sigma_{[M\}}^t$; thus we obtain $\Sigma_{L'}$.

Step 5:

$$\begin{aligned}
&\Sigma_{[M'\rangle}=R_{[M'\rangle}\Sigma_{L\cup M},\\
&\Sigma_{[M']}=\Lambda_{[M']}+R_{[M'\rangle}\Sigma_{\langle M']},\\
(2.76)\qquad&\Sigma_{[M'\}}=R_{[M'\rangle}\Sigma_{\langle M'\}}\\
&\left(=\Sigma_{[M'\rangle}\Sigma_{L\cup M}^{-1}\begin{pmatrix}\Sigma_{\langle L'']}\\ \Sigma_{\{L'']}\end{pmatrix}\right),
\end{aligned}$$

where $\Sigma_{\{L'']}=\Sigma_{[L''\}}^t$.

Thus $\mathbf{P}(\mathscr{K})$ consists of all Σ of the form

$$(2.77)\qquad \Sigma=\left(\begin{array}{cc:c:c:c}
\Sigma_{L\cap M} & \Sigma_{\langle L]} & \Sigma_{\langle M]} & \Sigma_{\langle L'']} & \Sigma_{\langle M']}\\
\Sigma_{[L\rangle} & \Sigma_{[L]} & \Sigma_{\{M]} & & \\
\hdashline
\Sigma_{[M\rangle} & \Sigma_{[M\}} & \Sigma_{[M]} & \Sigma_{\{L'']} & \\
\hdashline
\multicolumn{2}{c:}{\Sigma_{[L''\rangle}} & \Sigma_{[L''\}} & \Sigma_{[L'']} & \Sigma_{\{M']}\\
\hdashline
\multicolumn{3}{c:}{\Sigma_{[M'\rangle}} & \Sigma_{[M'\}} & \Sigma_{[M']}
\end{array}\right)$$

partitioned according to the ordered decomposition

$$(2.78)\qquad I=(L\cap M)\,\dot\cup\,[L]\,\dot\cup\,[M]\,\dot\cup\,[L'']\,\dot\cup\,[M'],$$

where $\Sigma_{[M\}}$, $\Sigma_{[L''\}}$ and $\Sigma_{[M'\}}$ satisfy (2.62), (2.75) and (2.76), respectively. The precision matrix $\Delta\equiv\Sigma^{-1}$ satisfies the following three conditions: Its $[M']\times[L'']$- and $[L'']\times[M']$-submatrices are 0, the $[L'']\times[M]$- and $[M]\times[L'']$-submatrices of $\Sigma_{L'}^{-1}$ are 0 and $\Sigma_{L\cup M}^{-1}$ has the form (2.65). Neither $\mathbf{P}(\mathscr{K})$ nor $\mathbf{P}(\mathscr{K})^{-1}$ is linear. The group $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I\times I$

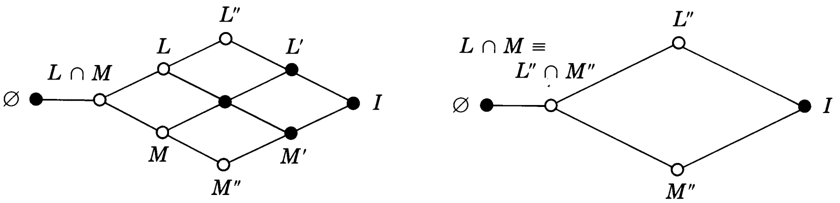

FIG. 9. *(a) The lattice $\mathscr{K}$. (b) The lattice $\mathscr{M}$.*

ces of the form

$$(2.79)\qquad A = \left(\begin{array}{cc:c:c:c} A_{L\cap M} & 0 & 0 & 0 & 0 \\ A_{[L\rangle} & A_{[L]} & 0 & 0 & 0 \\ \hdashline A_{[M\rangle} & 0 & A_{[M]} & 0 & 0 \\ \hdashline \multicolumn{2}{c:}{A_{[L''\rangle}} & 0 & A_{[L'']} & 0 \\ \hdashline \multicolumn{3}{c:}{A_{[M'\rangle}} & 0 & A_{[M']} \end{array}\right).$$

EXAMPLE 2.9. Finally, consider the lattice $\mathscr{K}$ in Figure 9a. Although this lattice properly contains the lattices in Examples 2.7 and 2.8 as sublattices, the set $\mathbf{P}(\mathscr{K})$ that it determines is much simpler than those in Examples 2.7 and 2.8. The reader may verify that $\mathbf{P}(\mathscr{K})$ is identical to $\mathbf{P}(\mathscr{M})$, where $\mathscr{M}$ is the sublattice in Figure 9b (compare to Example 2.5). Likewise, $\mathbf{P}(\mathscr{K}') = \mathbf{P}(\mathscr{M}')$ and $\mathbf{P}(\mathscr{K}'') = \mathbf{P}(\mathscr{M}'')$, where $\mathscr{K}'$ and $\mathscr{M}'$ are the sublattices in Figures 10a and b, respectively, and where $\mathscr{K}''$ and $\mathscr{M}''$ are the sublattices in Figures 11a and b, respectively.

For the original lattice $\mathscr{K}$, the group $\mathbf{GL}(\mathscr{K})$ consists of all nonsingular $I \times I$ matrices of the form

$$(2.80)\qquad A = \left(\begin{array}{c:c:c:c:c} A_{L\cap M} & 0 & 0 & 0 & 0 \\ \hdashline A_{[L\rangle} & A_{[L]} & 0 & 0 & 0 \\ \hdashline A_{[M\rangle} & 0 & A_{[M]} & 0 & 0 \\ \hdashline \multicolumn{2}{c:}{A_{[L''\rangle}} & 0 & A_{[L'']} & 0 \\ \hdashline A_{[M''(L\cap M)]} & 0 & A_{[M''M]} & 0 & A_{[M'']} \end{array}\right)$$

[Note that $(A_{[M''(L\cap M)]} \vdots A_{[M''M]}) = A_{[M''\rangle}$ in (2.80).]

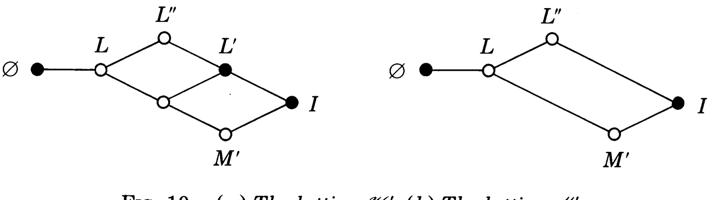

FIG. 10. *(a) The lattice $\mathscr{K}'$. (b) The lattice $\mathscr{M}'$.*

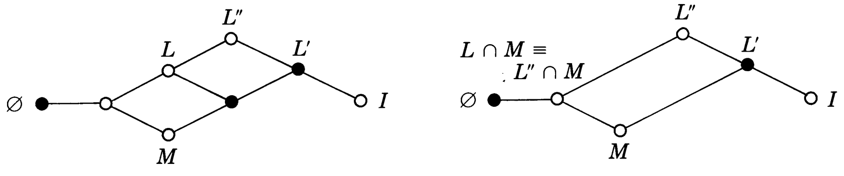

FIG. 11. *(a) The lattice $\mathscr{K}''$. (b) The lattice $\mathscr{M}''$.*

REMARK 2.5. For any $K \in \mathscr{D}(I)$ define $K' := I \setminus K$. It is an elementary exercise to verify that for $L, M \in \mathscr{D}(I)$, $x_L \perp\!\!\!\perp x_M | x_{L \cap M}$ under $N(\Sigma)$ if and only if $x_{L'} \perp\!\!\!\perp x_{M'} | x_{L' \cap M'}$ under $N(\Sigma^{-1})$. From this it follows that $\mathbf{P}(\mathscr{K}) = \mathbf{P}(\mathscr{K}')^{-1}$, where $\mathscr{K}' := \{K' | K \in \mathscr{K}\}$ is the dual lattice of $\mathscr{K}$. For example, if $\mathscr{K}$ is the lattice in Figure 4, then $\mathscr{K}'$ has the same form as the lattice in Figure 5; the relation $\mathbf{P}(\mathscr{K}) = \mathbf{P}(\mathscr{K}')^{-1}$ may be verified by comparing (2.57) and (2.65).

3. Likelihood inference for a normal model determined by pairwise conditional independence.

3.1. *Factorization of the likelihood function; the MLE of* Σ. Consider n independent, identically distributed (i.i.d.) observations $x_1, \dots, x_n$ from the lattice CI model $\mathbf{N}(\mathscr{K})$ defined by (1.8) and (1.4), and denote the matrix of observations by y, that is,

$$y := (x_1, \dots, x_n) \in \mathbf{M}(I \times N), \tag{3.1}$$

where $N = \{1, \dots, n\}$. For $L \in \mathscr{K}$ let y_L denote the $L \times N$ submatrix of y, while for $K \in J(\mathscr{K})$ partition y_K according to (2.3) as follows:

$$y_K = \begin{pmatrix} y_{\langle K \rangle} \\ y_{[K]} \end{pmatrix}.$$

The fundamental factorization of the LF for the model $\mathbf{N}(\mathscr{K})$ is an immediate consequence of Theorem 2.1(ii), Lemma 2.5 and Theorem 2.2.

THEOREM 3.1 (Factorization theorem). *The likelihood function based on n i.i.d. observations from the statistical model $\mathbf{N}(\mathscr{K})$ has the following factorization:*

$$\mathbf{P}(\mathscr{K}) \times \mathbf{M}(I \times N) \to]0, \infty[\,,$$

$$\begin{aligned} (\Sigma, y) &\to (\det(\Sigma))^{-n/2} \exp(-\operatorname{tr}(\Sigma^{-1} y y^t)/2) \\ &= \prod \Big((\det(\Sigma_{[K]\cdot}))^{-n/2} \\ &\quad \times \exp\big(-\operatorname{tr}(\Sigma_{[K]\cdot}^{-1} (y_{[K]} - \Sigma_{[K\rangle} \Sigma_{\langle K \rangle}^{-1} y_{\langle K \rangle})(\cdots)^t)/2\big) \Big| K \in J(\mathscr{K}) \Big). \end{aligned} \tag{3.2}$$

The parameter space $\mathbf{P}(\mathscr{K})$ has the factorization given by (2.32).

Note that the factor corresponding to $K \in J(\mathscr{K})$ is the density for the conditional distribution of $y_{[K]}$ given $y_{\langle K \rangle}$.

It follows readily from Theorem 3.1 and well-known results for the multivariate normal linear regression model that the MLE $\hat{\Sigma}(y)$ of $\Sigma \in \mathbf{P}(\mathscr{K})$ is unique if it exists, and it exists for a.e. $y \in \mathbf{M}(I \times N)$ if and only if

$$(3.3) \quad n \geq \max\{|\langle K \rangle| + |[K]| \,|K \in J(\mathscr{K})\} \equiv \max\{|K| \,|K \in J(\mathscr{K})\}.$$

In this case the $\mathscr{K}$-parameters of $\hat{\Sigma}$ are determined from the usual formulas for regression estimators:

$$(3.4) \qquad \widehat{\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}} = S_{[K\rangle}S_{\langle K\rangle}^{-1}, \qquad n\hat{\Sigma}_{[K]\cdot} = S_{[K]\cdot}, \qquad K \in J(\mathscr{K}),$$

where $S(y) = yy^t$ is n times the empirical covariance matrix. The explicit expression for $\hat{\Sigma}$ itself may be obtained from its $\mathscr{K}$-parameters in (3.4) by means of the reconstruction algorithm given in Section 2.7.

If $I \in J(\mathscr{K})$, then the condition (3.3) reduces to $N \geq |I|$, so in this case S is positive definite a.e., hence a fortiori $\hat{\Sigma}_{[K]\cdot}$ exists a.e. for every $K \in J(\mathscr{K})$. If, on the other hand, $I \notin J(\mathscr{K})$, then condition (3.3) does not guarantee that S is positive definite, but it still guarantees that $\hat{\Sigma}_{[K]\cdot}$ (and hence $\hat{\Sigma}$) exists a.e.

By Lemma 2.5, when (3.3) is satisfied the maximum value of the LF (3.2) is given by

$$(3.5) \qquad c \cdot \Pi\Big(\big(\det\big(\hat{\Sigma}_{[K]\cdot}\big)\big)^{-n/2} \Big| K \in J(\mathscr{K})\Big) = c \cdot \big(\det(\hat{\Sigma})\big)^{-n/2},$$

where $c = \exp(-n|I|/2)$. This fact is used in Andersson and Perlman (1993b) to express the likelihood ratio statistic for testing one model against another.

REMARK 3.1. The statistical model $\mathbf{N}(\mathscr{K})$ is a curved exponential family; it is linear if and only if $\mathbf{P}(\mathscr{K})^{-1}$ is a linear set, that is, closed under positive linear combinations. In the linear case the MLE $\hat{\Sigma}$ based on n i.i.d. observations from $\mathbf{N}(\mathscr{K})$ is a minimal sufficient statistic, but $\hat{\Sigma}$ is not necessarily sufficient in the general case.

3.2. *Examples of pairwise conditional independence models.* For each lattice $\mathscr{K}$ in Examples 2.1–2.9, consider the associated normal model $\mathbf{N}(\mathscr{K})$. When $\mathscr{K}$ is a chain as in Examples 2.1 and 2.2, $\mathbf{P}(\mathscr{K}) = \mathbf{P}(\mathrm{I})$ and $\mathbf{N}(\mathscr{K})$ is the unrestricted covariance model regardless of the length of the chain. [The $\mathscr{K}$-parametrization of $\mathbf{P}(\mathscr{K})$ does depend on this length, however.] Condition (3.3) for the existence of the MLE $\hat{\Sigma}$ reduces to the familiar condition $n \geq |I|$, while (3.4) reduces to $n\hat{\Sigma} = S$.

For the lattice $\mathscr{K}$ in Example 2.3, partition the observation $x \in \mathbb{R}^I$ according to (2.54) as $x = (x_L^t, x_M^t)^t$. The model $\mathbf{N}(\mathscr{K})$ states simply that $x_L \perp\!\!\!\perp x_M$. According to (3.3), the MLE $\hat{\Sigma}$ exists if and only if $n \geq \max\{|L|, |M|\}$ (whereas S is positive definite if and only if $n \geq |I|$) and is given by $n\hat{\Sigma} = \mathrm{Diag}(S_L, S_M)$.

For the lattice $\mathscr{K}$ in Example 2.4, partition $x \in \mathbb{R}^I$ according to (2.58) as $x = (x_L^t, x_M^t, x_{[I]}^t)^t$. Then the model $\mathbf{N}(\mathscr{K})$ again states that $x_L \perp\!\!\!\perp x_M$. Condition

(3.3) for the existence of the MLE takes the form $n \geq |I|$, while from (3.4).

$$n\hat{\Sigma}_L = S_L, \qquad n\hat{\Sigma}_M = S_M, \qquad \widehat{\Sigma_{[I\rangle}\Sigma^{-1}_{L\cup M}} = S_{[I\rangle}S^{-1}_{L\cup M}, \qquad n\hat{\Sigma}_{[I]\cdot} = S_{[I]\cdot}.$$

We reconstruct $\hat{\Sigma}$ from its $\mathscr{K}$-parameters by following steps 1–3 in Example 2.4 to obtain

$$n\hat{\Sigma}_{L\cup M} = \mathrm{Diag}(S_L, S_M),$$
$$n\hat{\Sigma}_{[I\rangle} = S_{[I\rangle}S^{-1}_{L\cup M}\,\mathrm{Diag}(S_L, S_M),$$
$$n\hat{\Sigma}_{[I]} = S_{[I]\cdot} + S_{[I\rangle}(\mathrm{Diag}(S_L, S_M))^{-1}S_{\langle I]}\ (\neq S_{[I]}).$$

In Example 2.5, x is partitioned according to (2.64) as $(x^t_{L\cap M}, x^t_{[L]}, x^t_{[M]})^t$. The model $\mathbf{N}(\mathscr{K})$ states that $x_{[L]} \perp\!\!\!\perp x_{[M]} | x_{L\cap M}$. Condition (3.3) becomes $n \geq \max\{|L|, |M|\}$, while (3.4) becomes

$$n\hat{\Sigma}_{L\cap M} = S_{L\cap M}, \tag{3.6a}$$

$$\widehat{\Sigma_{[L\rangle}\Sigma^{-1}_{L\cap M}} = S_{[L\rangle}S_{L\cap M}, \qquad n\hat{\Sigma}_{[L]\cdot} = S_{[L]\cdot}, \tag{3.6b}$$

$$\widehat{\Sigma_{[M\rangle}\Sigma^{-1}_{L\cap M}} = S_{[M\rangle}S_{L\cap M}, \qquad n\hat{\Sigma}_{[M]\cdot} = S_{[M]\cdot}. \tag{3.6c}$$

By steps 1–3 in Example 2.5, $\hat{\Sigma}$ is given by (3.6a) and

$$n\hat{\Sigma}_{[L\rangle} = S_{[L\rangle}, \qquad n\hat{\Sigma}_{[L]} = S_{[L]}, \tag{3.7a}$$

$$n\hat{\Sigma}_{[M\rangle} = S_{[M\rangle}, \qquad n\hat{\Sigma}_{[M]} = S_{[M]}, \tag{3.7b}$$

$$n\hat{\Sigma}_{[M\rangle\}} = S_{[M\rangle}S^{-1}_{L\cap M}S_{\langle L]}\ (\neq S_{[M\}}). \tag{3.7c}$$

In Example 2.6, x is partitioned as $(x^t_{L\cap M}, x^t_{[L]}, x^t_{[M]}, x^t_{[I]})^t$ and the model $\mathbf{N}(\mathscr{K})$ states that $x_{[L]} \perp\!\!\!\perp x_{[M]} | x_{L\cap M}$. Condition (3.3) reduces to $n \geq |I|$, while (3.4) is given by (3.6a)–(3.6c) and

$$\widehat{\Sigma_{[I\rangle}\Sigma^{-1}_{L\cup M}} = S_{[I\rangle}S^{-1}_{L\cup M}, \qquad n\hat{\Sigma}_{[I]\cdot} = S_{[I]\cdot}.$$

From steps 1–4 in Example 2.6, $\hat{\Sigma}$ is given by (3.6a) and (3.7a)–(3.7c) and

$$n\hat{\Sigma}_{[I\rangle} = S_{[I\rangle}S^{-1}_{L\cup M}\big(n\hat{\Sigma}_{L\cup M}\big)(\neq S_{[I\rangle})$$
$$n\hat{\Sigma}_{[I]} = S_{[I]\cdot} + S_{[I\rangle}S^{-1}_{L\cup M}\big(n\hat{\Sigma}_{L\cup M}\big)S^{-1}_{L\cup M}S_{\langle I]}\ (\neq S_{[I]}),$$

where, from (3.6a) and (3.7a)–(3.7c),

$$n\hat{\Sigma}_{L\cup M} = \begin{pmatrix} S_{L\cap M} & S_{\langle L]} & S_{\langle M]} \\ S_{[L\rangle} & S_{[L]} & S_{[L\rangle}S^{-1}_{L\cap M}S_{\langle M]} \\ S_{[M\rangle} & S_{[M\rangle}S^{-1}_{L\cap M}S_{\langle L]} & S_{[M]} \end{pmatrix}. \tag{3.8}$$

In Example 2.7, x is partitioned as $(x_{L\cap M}^t, x_{[L]}^t, x_{[M]}^t, x_{[L']}^t, x_{[M']}^t)^t$ and the model $\mathbf{N}(\mathscr{K})$ states that

$$x_{[L]} \perp\!\!\!\perp x_{[M]} | x_{L\cap M} \quad \text{and that} \quad x_{[L']} \perp\!\!\!\perp x_{[M']} | (x_{L\cap M}, x_{[L]}, x_{[M]}).$$

[Note that $x_{L'\cap M'} = x_{L\cup M} = (x_{L\cap M}, x_{[L]}, x_{[M]})$.] Condition (3.3) becomes $n \geq \max\{|L'|, |M'|\}$, while (3.4) is given by (3.6a)–(3.6c) and by (3.6b) and (3.6c) with L, M replaced by L', M' (note that $S_{L'\cap M'} = S_{L\cup M}$). From steps 1–5 in Example 2.7, $\hat{\Sigma}$ is given by (3.6a), (3.7a)–(3.7c) and

$$n\hat{\Sigma}_{[L'\rangle} = S_{[L'\rangle} S_{L\cup M}^{-1}\big(n\hat{\Sigma}_{L\cup M}\big),$$

$$n\hat{\Sigma}_{[L']} = S_{[L']\cdot} + S_{[L'\rangle} S_{L\cup M}^{-1}\big(n\hat{\Sigma}_{L\cup M}\big) S_{L\cup M}^{-1} S_{\langle L']},$$

(3.9a) $$n\hat{\Sigma}_{[M'\rangle} = S_{[M'\rangle} S_{L\cup M}^{-1}\big(n\hat{\Sigma}_{L\cup M}\big),$$

(3.9b) $$n\hat{\Sigma}_{[M']} = S_{[M']\cdot} + S_{[M'\rangle} S_{L\cup M}^{-1}\big(n\hat{\Sigma}_{L\cup M}\big) S_{L\cup M}^{-1} S_{\langle M']},$$

(3.9c) $$n\hat{\Sigma}_{[M'\rbrace} = S_{[M'\rangle} S_{L\cup M}^{-1}\big(n\hat{\Sigma}_{L\cup M}\big) S_{L\cup M}^{-1} S_{\langle L']} \ (\neq S_{[M'\rbrace}),$$

where $n\hat{\Sigma}_{L\cup M}$ is given by (3.8).

In Example 2.8, x is partitioned as $(x_{L\cap M}^t, x_{[L]}^t, x_{[M]}^t, x_{[L'']}^t, x_{[M']}^t)^t$. It may be seen from the form (2.77) of $\Sigma \in \mathbf{P}(\mathscr{K})$ [or from (3.11)] that the model $\mathbf{N}(\mathscr{K})$ is determined by the following three conditions:

(i) $$x_{[L]} \perp\!\!\!\perp x_{[M]} | x_{L\cap M},$$

(ii) $$x_{[M]} \perp\!\!\!\perp x_{[L'']} | (x_{L\cap M}, x_{[L]}),$$

(iii) $$x_{[L'']} \perp\!\!\!\perp x_{[M']} | (x_{L\cap M}, x_{[L]}, x_{[M]}).$$

Condition (3.3) becomes $n \geq \max\{|L''|, |M'|\}$, while (3.4) is given by (3.6a)–(3.6c),

$$\widehat{\Sigma_{[L''\rangle}\Sigma_L^{-1}} = S_{[L''\rangle} S_L^{-1}, \qquad n\hat{\Sigma}_{[L'']\cdot} = S_{[L'']\cdot},$$

$$\widehat{\Sigma_{[M'\rangle}\Sigma_{L\cup M}^{-1}} = S_{[M'\rangle} S_{L\cup M}^{-1}, \qquad n\hat{\Sigma}_{[M']\cdot} = S_{[M']\cdot}.$$

From steps 1–5 in Example 2.8, $\hat{\Sigma}$ is given by (3.9a) and (3.7a)–(3.7c), by

$$n\hat{\Sigma}_{[L''\rangle} = S_{[L''\rangle}, \qquad n\hat{\Sigma}_{[L'']} = S_{[L'']},$$

$$n\hat{\Sigma}_{[L''\rbrace} = S_{[L''\rangle} S_L^{-1} \begin{pmatrix} S_{\langle M]} \\ S_{[L\rangle} S_{L\cap M}^{-1} S_{\langle M]} \end{pmatrix},$$

by (3.9a) and (3.9b), and by

$$n\hat{\Sigma}_{[M'\rbrace} = S_{[M'\rangle} S_{L\cup M}^{-1} \begin{pmatrix} S_{\langle L'']} \\ n\hat{\Sigma}_{\lbrace L'']} \end{pmatrix}.$$

Finally, for the lattice $\mathscr{K}$ in Example 2.9, x is partitioned as

$$\left(x_{L\cap M}^t, x_{[L]}^t, x_{[M]}^t, x_{[L'']}^t, x_{[M'']}^t\right)^t.$$

It is readily seen (cf. Remark 3.2) that the model $\mathbf{N}(\mathscr{K})$ is determined by the single condition that

$$(x_{[L]}, x_{[L'']}) \perp\!\!\!\perp (x_{[M]}, x_{[M'']})|x_{L\cap M}.$$

This reflects the fact that this model is of the same form as that in Example 2.5 (see the discussion in Example 2.9).

REMARK 3.2. The normal model $\mathbf{N}(\mathscr{K})$ is defined in terms of the set of *all* CI conditions determined by $\mathscr{K}$:

$$(3.10)\qquad \Sigma \in \mathbf{P}(\mathscr{K}) \Leftrightarrow x_L \perp\!\!\!\perp x_M | x_{L\cap M} \qquad \forall\, L, M \in \mathscr{K}.$$

It may be seen from the previous examples that many of these conditions are trivial (e.g., when $L \subseteq M$ or $M \subseteq L$) or redundant and may be omitted. (For example, the CI condition is redundant when $L \subseteq L'$, $M \subseteq M'$ and $L \cap M = L' \cap M'$, for then $x_{L'} \perp\!\!\!\perp x_{M'}|x_{L'\cap M'} \Rightarrow x_L \perp\!\!\!\perp x_M|x_{L\cap M}$.) Any never-decreasing listing $K_1, \dots, K_q$ of the members of $J(\mathscr{K})$ (cf. Section 2.7) may be used to replace the set (3.10) of CI conditions by a subset (3.11) of only $q-1$ CI conditions (some possibly trivial) that determines the same model $\mathbf{N}(\mathscr{K})$:

$$(3.11)\quad \Sigma \in \mathbf{P}(\mathscr{K}) \Leftrightarrow H_k\colon x_k \perp\!\!\!\perp x_{1\cup\cdots\cup(k-1)}|x_{\langle k\rangle}, \qquad k = 2,\dots,q,$$

$$\Leftrightarrow H_k\colon x_{[k]} \perp\!\!\!\perp x_{(1\cup\cdots\cup(k-1))\setminus\langle k\rangle}|x_{\langle k\rangle}, \; k = 2,\dots,q,$$

$$\Leftrightarrow H_k\colon \Sigma_{[k-\Sigma[k\rangle}\Sigma_{\langle k\rangle}^{-1}\Sigma_{\langle k\}} = 0, \qquad k = 2,\dots,q.$$

Here we use the notational conventions of Section 2.7 and the equality in (2.42). The equivalence of (3.10) and (3.11) follows from the fact that (3.11) implies the factorization (3.2), which in turn implies that condition (ii) of Theorem 2.1 holds. [We thank the referee for suggesting condition (3.11).]

In Example 2.8 each of the following four sets of CI conditions determines $\mathbf{N}(\mathscr{K})$:

(i) $x_L \perp\!\!\!\perp x_M|x_{L\cap M}$; (ii) $x_{L\cup M} \perp\!\!\!\perp x_{L''}|x_L$; (iii) $x_{L'} \perp\!\!\!\perp x_{M'}|x_{L\cup M}$;

(i) $x_L \perp\!\!\!\perp x_M|x_{L\cap M}$; (ii) $x_{L''} \perp\!\!\!\perp x_{M'}|x_L$;

(i) $x_M \perp\!\!\!\perp x_{L''}|x_{L\cap M}$; (ii) $x_{L'} \perp\!\!\!\perp x_{M'}|x_{L\cup M}$;

(i) $x_M \perp\!\!\!\perp x_{L''}|x_{L\cap M}$; (ii) $x_{L''} \perp\!\!\!\perp x_{M'}|x_L$.

The first three sets are of the form (3.11) (omitting trivial conditions), corresponding to the three never-decreasing listings $(L\cap M, L, M, L'', M')$, $(L\cap M, L, M, M', L'')$ and $(L\cap M, L, L'', M, M')$ of $J(\mathscr{K})$. The fourth, however, is not of the form (3.11).

REMARK 3.3. For $I = \{1, 2, 3, 4\}$, consider the statistical model consisting of all normal distributions on $\mathbb{R}^I$ such that x_1 is independent of x_2 and x_3 is independent of x_4. It is readily seen that this model is *not* of the form $\mathbf{N}(\mathscr{K})$ for any $\mathscr{K}$. The same is true for the normal model determined by the two conditions that x_1 and x_2 are CI given (x_3, x_4) and x_3 and x_4 are CI given (x_1, x_2).

REMARK 3.4. The general model $\mathbf{N}(\mathscr{K})$ is defined by the *pairwise* CI requirement (1.4) for every pair $L, M \in \mathscr{K}$. This requirement does not necessarily imply, however, that for every subset $\mathscr{S} \subseteq \mathscr{K}$, $(x_K | K \in \mathscr{S})$ are mutually CI given $x_{\cap(K|K \in \mathscr{S})}$. For the lattice $\mathscr{K}$ in Example 2.9, this may be seen by considering the subset $\mathscr{S} = \{L'', L \cup M, M''\}$.

REMARK 3.5. An alternative statistical interpretation of the CI model $\mathbf{N}(\mathscr{K})$ may be obtained from (2.35): $x \equiv (x_{[K]} | K \in J(\mathscr{K})) \in \mathbb{R}^I$ is an observation from the normal model $\mathbf{N}(\mathscr{K})$ if and only if x can be represented in the form $x = Az$ for some (generalized block-triangular) matrix $A \in \mathbf{GL}(\mathscr{K})$, where $z \equiv (z_{[K]} | K \in J(\mathscr{K})) \in \mathbb{R}^I$ is an (unobservable) stochastic variate such that $z \sim N(1_I)$. From Proposition 2.2(iii), this representation is equivalent to the system of equations

$$x_{[L]} = \Sigma\big(A_{[LM]} z_{[M]} | M \in H(L)\big), \qquad L \in J(\mathscr{K}), \tag{3.12}$$

where $H(L) := \{M \in J(\mathscr{K}) | M \subseteq L\} \equiv J(\mathscr{K}_L)$. Since $A \in \mathbf{GL}(\mathscr{K}) \Rightarrow A^{-1} \in \mathbf{GL}(\mathscr{K})$, this shows that the CI model $\mathbf{N}(\mathscr{K})$ can be interpreted as a multivariate linear recursive model [cf. Wermuth (1980) and Kiiveri, Speed and Carlin (1984)] with additional lattice constraints.

Conversely, suppose that J is a finite index set and let $(H(\ell) | \ell \in J)$ be a family of subsets of J that satisfies the following two conditions:

(i) $$\ell \in H(\ell),$$

(ii) $$m \in H(\ell) \Rightarrow H(m) \subseteq H(\ell).$$

For each $\ell \in J$ let D_ℓ and E_ℓ be finite index sets such that $|D_\ell| \le |E_\ell|$ and let $I = \dot{\bigcup}(D_\ell | \ell \in J)$, $I' = \dot{\bigcup}(E_\ell | \ell \in J)$. Consider the normal statistical model defined by the system of equations

$$x_{[\ell]} = \Sigma\big(A_{\ell m} z_{[m]} | m \in H(\ell)\big), \qquad \ell \in J, \tag{3.13}$$

where $x_{[\ell]} \in \mathbb{R}^{D_\ell}$ is observable, $z_{[m]} \in \mathbb{R}^{E_m}$ is unobservable, $z \equiv (z_{[m]} | m \in J) \sim N(1_{I'})$ on $\mathbb{R}^{I'}$, $A_{\ell m} \in \mathbf{M}(D_\ell \times E_m)$, and $\operatorname{rank}(A_{\ell\ell}) = |D_\ell|$. Let $\mathscr{H}$ be the ring of subsets of J generated by $\{H(\ell) | \ell \in J\}$ and for $H \in \mathscr{H}$ define $I_H = \dot{\bigcup}(D_\ell | \ell \in H)$. Then trivially $\mathscr{K} := \{I_H | H \in \mathscr{H}\}$ is a ring of subsets of I and the model determined by the system (3.13) has the form (3.12), that is, it is the model $\mathbf{N}(\mathscr{K})$.

3.3. *Invariance of the model.* It follows from the well-known transformation property of the multivariate normal distribution that the i.i.d. model determined by $\mathbf{N}(\mathscr{K})$ is invariant under the transitive action (2.33) of $\mathbf{GL}(\mathscr{K})$ on the parameter space $\mathbf{P}(\mathscr{K})$ and the action

$$\begin{aligned}(3.14)\qquad \mathbf{GL}(\mathscr{K}) \times \mathbf{M}(I \times N) &\to \mathbf{M}(I \times N),\\ (A, y) &\to Ay\end{aligned}$$

of $\mathbf{GL}(\mathscr{K})$ on the observation space $\mathbf{M}(I \times N)$. The MLE is thus equivariant.

4. Invariant formulation of the CI model and testing problem.

4.1. *The lattice structure of quotient spaces.* Let V be a finite-dimensional real vector space. A *quotient space* (or simply a *quotient*) of V is formally defined to be a pair (Q, p_Q) consisting of a vector space Q and a surjective linear mapping $p_Q\colon V \to Q$. For ease of notation, (Q, p_Q) usually is abbreviated to Q.

Let R and T be two quotients of V. If there exists a linear mapping $p_{RT}\colon T \to R$ such that $p_R = p_{RT} \circ p_T$, then p_{RT} is necessarily surjective and unique; hence (R, p_{RT}) is a quotient of T. In this situation we write $(R, p_R) \le (T, p_T)$, or simply $R \le T$. This relation is equivalent to the condition that $p_R^{-1}(0) \supseteq p_T^{-1}(0)$. The relation $\le$ on the set of all quotients of V is not antisymmetric; hence one defines an equivalence relation $\sim$ on this set by $R \sim T$ if $p_R^{-1}(0) = p_T^{-1}(0)$. The collection of equivalence classes is denoted by $\mathscr{Q}(V)$. Equipped with the relation induced by $\le$ (also denoted by $\le$), $\mathscr{Q}(V)$ becomes a partially ordered set ($\equiv$ *poset*).

We identify a quotient (Q, p_Q) of V with its equivalence class in $\mathscr{Q}(V)$. A convenient representative for this equivalence class is the canonical quotient space $(V/p_Q^{-1}(0), p)$, where $p\colon V \to V/p_Q^{-1}(0)$ is the canonical quotient mapping given by $p(x) = x + p_Q^{-1}(0)$, $x \in V$.

The poset $\mathscr{Q}(V)$ is in fact a lattice: If $R, T \in \mathscr{Q}(V)$, then their minimum and maximum exist and are given by

$$R \wedge T := V/\left(p_R^{-1}(0) + p_T^{-1}(0)\right),$$

$$R \vee T := V/\left(p_R^{-1}(0) \cap p_T^{-1}(0)\right),$$

respectively. The minimal and maximal elements exist and are given by $\{0\}$ and V, respectively. If $\dim(V) \ge 2$, then $\mathscr{Q}(V)$ is not distributive and $|\mathscr{Q}(V)| = \infty$. Since V is finite dimensional, the lattice $\mathscr{Q}(V)$ has finite length; hence so does any sublattice $\mathscr{Q} \subseteq \mathscr{Q}(V)$. Therefore, if $\mathscr{Q}$ is a distributive sublattice of $\mathscr{Q}(V)$, it must be finite. The reader is referred to Section 3 of Andersson (1990) for the properties of posets and lattices used here.

4.2. *Invariant formulation of the pairwise CI model.* For $\sigma \in \mathbf{P}(V) :=$ the cone of all positive definite forms on the dual vector space V^* of V, let $N(\sigma)$ denote the normal distribution on V with mean vector $0 \in V$ and covariance σ [cf. Andersson (1975), Section 5]. Let $\mathscr{Q} \subseteq \mathscr{Q}(V)$ be a sublattice such that $\{0\}, V \in \mathscr{Q}$.

DEFINITION 4.1. The class $\mathbf{P}(\mathscr{D}) \subseteq \mathbf{P}(V)$ is defined as follows:

$$\sigma \in \mathbf{P}(\mathscr{D}) \Leftrightarrow p_R(x) \perp\!\!\!\perp p_T(x) | p_{R \wedge T}(x) \quad \forall R, T \in \mathscr{D} \text{ when } x \sim N(\sigma), \tag{4.1}$$

that is, p_R and p_T are conditionally independent (CI) given $p_{R \wedge T}$ (compare to Definition 2.1).

THEOREM 4.1. *The class* $\mathbf{P}(\mathscr{D})$ *is nonempty if and only if the lattice* $\mathscr{D}$ *is distributive.*

PROOF. See Section A.2.

The normal statistical model $\mathbf{N}(\mathscr{D})$ defined by the requirement (4.1) of pairwise conditional independence wrt $\mathscr{D}$ is then given by

$$\mathbf{N}(\mathscr{D}) := (N(\sigma) | \sigma \in \mathbf{P}(\mathscr{D})) \tag{4.2}$$

[compare to (1.8)]. By Theorem 4.1, $\mathbf{N}(\mathscr{D}) \neq \varnothing$ if and only if $\mathscr{D}$ is distributive.

EXAMPLE 4.1. Let $V = \mathbb{R}^I$, where I is a finite index set. Every subring $\mathscr{K} \subseteq \mathscr{D}(I)$ determines a distributive sublattice $\mathscr{D}(\mathscr{K}) \subseteq \mathscr{D}(\mathbb{R}^I)$ as follows. For each $K \in \mathscr{K}$ define the coordinate projection p_K: $\mathbb{R}^I \to \mathbb{R}^K$ by $p_K((x_i | i \in I)) = (x_i | i \in K)$. Since $\mathscr{K}$ is a ring, it follows that $\mathscr{D}(\mathscr{K}) := \{(\mathbb{R}^K, p_K) | K \in \mathscr{K}\}$ is a distributive lattice of quotients of $\mathbb{R}^I$. If $\varnothing, I \in \mathscr{K}$, then $\{0\}, \mathbb{R}^I \in \mathscr{D}(\mathscr{K})$. Thus each canonical coordinate-wise CI model $\mathbf{N}(\mathscr{K})$ given by (1.8) is a special case of the general CI model $\mathbf{N}(\mathscr{D})$ given by (4.2).

Conversely, by Proposition 4.1 *every* distributive sublattice $\mathscr{D} \subseteq \mathscr{D}(V)$ can be represented in the form $\mathscr{D} = \mathscr{D}(\mathscr{K})$ for some ring of subsets $\mathscr{K}$ and *every* CI model $\mathbf{N}(\mathscr{D})$ can be represented as a canonical model $\mathbf{N}(\mathscr{K})$.

4.3. *Reduction of the CI model to canonical coordinate-wise form.*

PROPOSITION 4.1. *Let* $\mathscr{D} \subseteq \mathscr{D}(V)$ *be a distributive lattice of quotients. Then there exists a set* I, *a ring* $\mathscr{K}$ *of subsets of* I *with the property* $\varnothing, I \in \mathscr{K}$, *a lattice isomorphism* $Q \to K(Q)$ *of* $\mathscr{D} \to \mathscr{K}$ *and a basis* $(e_i | i \in I)$ *for* V *such that the quotients* $(Q, P_Q) \in \mathscr{D}$ *can be represented as follows*:

$$Q = \operatorname{span}\{e_i | i \in K(Q)\}, \tag{4.3}$$

$$P_Q(e_i) = \begin{cases} e_i, & \text{for } i \in K(Q), \\ 0, & \text{for } i \in I \setminus K(Q). \end{cases} \tag{4.4}$$

PROOF. See Section A.1.

We say that a basis $(e_i | i \in I)$ for V satisfying the conditions in Proposition 4.1 is *adapted* to $\mathscr{D}$. Thus, when V is identified with $\mathbb{R}^I$ through a $\mathscr{D}$-adapted basis $(e_i | i \in I)$, the distributive lattice $\mathscr{D} \subseteq \mathscr{D}(V)$ is identified with the ring $\mathscr{K}(\mathscr{D}) := \{K(Q) | Q \in \mathscr{D}\}$ of subsets of I and the quotients P_Q, $Q \in \mathscr{D}$, are identified with the coordinate projections $p_K: \mathbb{R}^I \to \mathbb{R}^K$, $K \in \mathscr{K}(\mathscr{D})$ (cf. Example 4.1). Furthermore, $\mathbf{P}(V)$ is identified with $\mathbf{P}(I)$ through the correspondence $\sigma \to \Sigma$, where Σ is the matrix of σ wrt the dual basis $(e_i^* | i \in I)$ for V^*. The condition (4.1) is then transformed into the condition (2.1); hence $\mathbf{P}(\mathscr{D})$ is identified with $\mathbf{P}(\mathscr{K}(\mathscr{D}))$ and the model $\mathbf{N}(\mathscr{D})$ is transformed into the canonical form $\mathbf{N}(\mathscr{K}(\mathscr{D}))$.

REMARK 4.1. Since the identity matrix $1_I \in \mathbf{P}(\mathscr{K}(\mathscr{D}))$, the model $\mathbf{N}(\mathscr{D})$ is nonempty when $\mathscr{D}$ is distributive.

4.4. *Invariant formulation of the testing problem.* Let $\mathscr{D}$ and $\mathscr{T}$ be two distributive sublattices of $\mathscr{D}(V)$ such that $\mathscr{T} \subset \mathscr{D}$. Then $\mathbf{P}(\mathscr{D}) \subseteq \mathbf{P}(\mathscr{T})$ and one may consider the general problem of testing $\mathbf{N}(\mathscr{D})$ against the (possible) larger model $\mathbf{N}(\mathscr{T})$ on the basis of n i.i.d. observations from V, that is, testing

$$H_{\mathscr{D}}: \sigma \in \mathbf{P}(\mathscr{D}) \quad \text{vs.} \quad H_{\mathscr{T}}: \sigma \in \mathbf{P}(\mathscr{T}). \tag{4.5}$$

By Proposition 4.1 we may choose a $\mathscr{D}$-adapted basis $(e_i | i \in I)$ for V; clearly this basis is also adapted to $\mathscr{T}$. It follows immediately that the testing problem (4.5) is transformed into the canonical testing problem (1.9) by this choice of a $\mathscr{D}$-adapted basis.

APPENDIX

In Sections A.1 and A.2, the notation and terminology of Section 4 are followed.

A.1. The decomposition theorem and existence of a $\mathscr{K}$-adapted basis.

LEMMA A.1. *For $R \in \mathscr{D}(V)$, the set $\mathscr{D}(V)_R := \{Q \in \mathscr{D} | Q \le R\}$ is a sublattice of $\mathscr{D}(V)$ isomorphic to the lattice $\mathscr{D}(R)$ of quotients of R through the lattice isomorphism*

$$\mathscr{D}(R) \leftrightarrow \mathscr{D}(V)_R,$$
$$(Q, p_{QR}) \leftrightarrow (Q, p_{QR} \circ p_R).$$

PROOF. Straightforward. □

LEMMA A.2. *Let $R, T \in \mathscr{D}(V)$ with $R \vee T = V$ and let $r_R: R \to p_{R \wedge T, R}^{-1}(0)$ and $r_T: T \to p_{R \wedge T, T}^{-1}(0)$ be surjective linear mappings. Then the linear mapping*

$$\begin{aligned} \varphi: V &\to (R \wedge T) \times p_{R \wedge T, R}^{-1}(0) \times p_{R \wedge T, T}^{-1}(0), \\ x &\to (p_{R \wedge T}(x), r_R(p_R(x)), r_T(p_T(x))) \end{aligned} \tag{A.1}$$

is bijective.

PROOF. Suppose that $\varphi(x) = 0$. Then $p_{R \wedge T}(x) = 0$ and we obtain that $p_R(x) \in p_{R \wedge T, R}^{-1}(0)$. In fact, $p_R(x) = 0$ since r_R is surjective. Similarly, $p_T(x) = 0$; hence $x \in p_R^{-1}(0) \cap p_T^{-1}(0) = \{0\}$. The linear mapping φ is thus injective. Since $\dim(V) = \dim((R \wedge T) \times p_{R \wedge T, R}^{-1}(0) \times p_{R \wedge T, T}^{-1}(0))$, φ is also surjective. □

As in Section 4.3 let $\mathscr{Q}$ be a distributive sublattice of $\mathscr{Q}(V)$ such that $\{0\}$, $V \in \mathscr{Q}(V)$. For $Q \in \mathscr{Q}$, $Q \neq \{0\}$, define

$$\langle Q \rangle := \bigvee (Q' \in \mathscr{Q} | Q' < Q)$$

and let $J(\mathscr{Q})$ denote the poset of all join-irreducible elements in $\mathscr{Q}$, that is,

$$\begin{aligned} J(\mathscr{Q}) &:= \{Q \in \mathscr{Q} | Q \neq \{0\}, \langle Q \rangle < Q\} \\ &= \{Q \in \mathscr{Q} | Q \neq \{0\}, \forall R, T \in \mathscr{Q}: Q = R \vee T \Rightarrow Q = R \text{ or } Q = T\}. \end{aligned}$$

In the following theorem the space V is represented as a product of vector spaces indexed by $J(\mathscr{Q})$ such that the space with index $Q \in J(\mathscr{Q})$ has dimension $\dim(Q) - \dim(\langle Q \rangle)$.

THEOREM A.1 (Decomposition theorem). *For each $Q \in J(\mathscr{Q})$, let r_Q: $Q \to p_{\langle Q \rangle, Q}^{-1}(0)$ be any surjective linear mapping. Then the linear mapping*

$$\text{(A.2)} \qquad \begin{aligned} \varphi_V: V &\to \times \left(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{Q})\right), \\ x &\to \left(r_Q(p_Q(x)) | Q \in J(\mathscr{Q})\right) \end{aligned}$$

is bijective.

PROOF. For $R \in \mathscr{Q}$ define $\mathscr{Q}_R := \{Q \in \mathscr{Q} | Q \leq R\}$, a sublattice of $\mathscr{Q}$ ($\mathscr{Q}_V \equiv \mathscr{Q}$). Then

$$\text{(A.3)} \qquad R = \bigvee (Q \in J(\mathscr{Q}_R)),$$

$$\text{(A.4)} \qquad J(\mathscr{Q}_R) = J(\mathscr{Q}) \cap \mathscr{Q}_R,$$

$$\text{(A.5)} \qquad J(\mathscr{Q}_{R \wedge T}) = J(\mathscr{Q}_R) \cap J(\mathscr{Q}_T),$$

$$\text{(A.6)} \qquad J(\mathscr{Q}_{R \vee T}) = J(\mathscr{Q}_R) \cup J(\mathscr{Q}_T)$$

[cf. (2.4)–(2.7)]. The proof proceeds by induction on $|J(\mathscr{Q})| =: q$. If $q = 1$, then $\mathscr{Q} = \{\{0\}, V\}$ and the result is trivial. Next, assume that the result is true whenever $q \leq k - 1$ and suppose that $q = k$. If $V \in J(\mathscr{Q})$, then $|J(\mathscr{Q}_{\langle V \rangle})| = k - 1$; hence the mapping

$$\begin{aligned} \langle V \rangle &\to \times \left(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{Q}_{\langle V \rangle})\right), \\ x &\to \left(r_Q(p_{Q, \langle V \rangle}(x)) | Q \in J(\mathscr{Q}_{\langle V \rangle})\right) \end{aligned}$$

is bijective by the induction assumption and Lemma A.1. Since the linear mapping

$$V \to \langle V \rangle \times p_{\langle V \rangle}^{-1}(0),$$
$$x \to \big(p_{\langle V \rangle}(x), r_V(x)\big)$$

is bijective and $p_{Q,\langle V \rangle} \circ p_{\langle V \rangle} = p_Q$ for every $Q \in J(\mathscr{D}_{\langle V \rangle})$, the mapping (A.2) is bijective in this case.

If, on the other hand, $V \notin J(\mathscr{D})$, then $V = R \vee T$ where $R < V$ and $T < V$. It follows from (A.3) that $|J(\mathscr{D}_R)| < k$ and $|J(\mathscr{D}_T)| < k$, so by the induction assumption and Lemma A.1, the mapping

$$V \to \times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_R)\big),$$
$$x \to \big(r_Q(p_Q(x)) | Q \in J(\mathscr{D}_R)\big)$$

is (equivalent to) the quotient mapping p_R: $V \to R$. Similarly, the quotient mappings p_T and $p_{R \wedge T}$ can be represented in an analogous way; hence

$$p_{R \wedge T, R}^{-1}(0) = \times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_R) \setminus J(\mathscr{D}_{R \wedge T})\big),$$
$$p_{R \wedge T, T}^{-1}(0) = \times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_T) \setminus J(\mathscr{D}_{R \wedge T})\big).$$

Thus, by (A.5) and (A.6),

$$\begin{aligned}\times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D})\big) = {} & \times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_{R \wedge T})\big) \\ & \times \times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_R) \setminus J(\mathscr{D}_{R \wedge T})\big) \\ & \times \times\big(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_T) \setminus J(\mathscr{D}_{R \wedge T})\big).\end{aligned}$$

Lemma A.2 now implies that φ_V is bijective. □

REMARK A.1. The representation (A.2) shows that V can be identified with a product of vector spaces indexed by $J(\mathscr{D})$; similarly, each $R \in \mathscr{D}$ can be identified with the product $\times(p_{\langle Q \rangle, Q}^{-1}(0) | Q \in J(\mathscr{D}_R))$ through the bijective linear mapping φ_R defined by $\varphi_R(x) = (r_Q(p_Q(x)) | Q \in J(\mathscr{D}_R))$, $x \in R$; under these identifications, each mapping p_{RT}, $R \le T \le V$, is simply a canonical projection mapping.

PROOF OF PROPOSITION 4.1. For each $Q \in J(\mathscr{D})$, let $[K(Q)]$ be a set with

$$|[K(Q)]| = \dim\big(p_{\langle Q \rangle, Q}^{-1}(0)\big).$$

For $R \in \mathscr{D}$, define

$$K(R) := \dot{\bigcup}\,([K(Q)] | Q \in J(\mathscr{D}_R)) \tag{A.7}$$

and define $I := K(V)$. From (A.5) and (A.6) it follows that $\mathscr{K} \equiv \mathscr{K}(\mathscr{D}) := \{K(R) | R \in \mathscr{D}\}$ is a subring of $\mathscr{D}(I)$ and the mapping $R \to K(R)$ is a lattice isomorphism between $\mathscr{D}$ and $\mathscr{K}$. Now Remark A.1 implies that there exists a basis $(e_i|i \in I)$ for V such that the elements (R, p_R) in $\mathscr{Q}$ can be represented as in (4.3) and (4.4). □

A.2. Proof of Theorem 4.1.

LEMMA A.3. *Suppose that $x \sim N(\sigma)$, $\sigma \in \mathbf{P}(V)$. Then for any $R, T \in \mathscr{Q}(V)$, $p_R(x) \perp\!\!\!\perp p_T(x) | p_{R \wedge T}(x) \Leftrightarrow p_R^{-1}(0)$ and $p_T^{-1}(0)$ are geometrically orthogonal (g.o.) wrt the inner product $\delta := \sigma^{-1}$ on V [cf. Andersson (1990), Definition 4.1, for the definition of g.o.].*

PROOF. Let $p_R^{-1}(0)^\perp$, $p_T^{-1}(0)^\perp$ and $p_{R \wedge T}^{-1}(0)^\perp$ denote the orthogonal complements of $p_R^{-1}(0)$, $p_T^{-1}(0)$ and $p_{R \wedge T}^{-1}(0)$, respectively, wrt δ. Furthermore, let q_R, q_T and $q_{R \wedge T}$ be the orthogonal projections of V onto $p_R^{-1}(0)^\perp$, $p_T^{-1}(0)^\perp$ and $p_{R \wedge T}^{-1}(0)^\perp$. Then $(p_R^{-1}(0)^\perp, q_R)$, $(p_T^{-1}(0)^\perp, q_T)$ and $(p_{R \wedge T}^{-1}(0)^\perp, q_{R \wedge T})$ represent the quotients (R, p_R), (T, p_T) and $(R \wedge T, p_{R \wedge T})$, respectively. Therefore,

$$
\begin{aligned}
& p_R(x) \perp\!\!\!\perp p_T(x) | p_{R \wedge T}(x) \\
& \quad \Leftrightarrow q_R(x) \perp\!\!\!\perp q_T(x) | q_{R \wedge T}(x) \\
& \quad \Leftrightarrow (q_R(x) - q_{R \wedge T}(x)) \perp\!\!\!\perp (q_T(x) - q_{R \wedge T}(x)) | q_{R \wedge T}(x) \\
& \quad \Leftrightarrow (q_R(x) - q_{R \wedge T}(x)) \perp\!\!\!\perp (q_T(x) - q_{R \wedge T}(x)). \\
& \quad \Leftrightarrow \left(p_R^{-1}(0)^\perp \cap p_{R \wedge T}^{-1}(0)\right) \perp \left(p_T^{-1}(0)^\perp \cap p_{R \wedge T}^{-1}(0)\right) \\
& \quad \Leftrightarrow p_R^{-1}(0)^\perp \text{ and } p_T^{-1}(0)^\perp \text{ are g.o.} \\
& \quad \Leftrightarrow p_R^{-1}(0) \text{ and } p_T^{-1}(0) \text{ are g.o.}
\end{aligned}
$$

The third $\Leftrightarrow$ follows since $(q_R - q_{R \wedge T}, q_T - q_{R \wedge T})$ is a projection onto $(p_R^{-1}(0)^\perp \cap p_{R \wedge T}^{-1}(0)) \oplus (p_T^{-1}(0)^\perp \cap p_{R \wedge T}^{-1}(0))$, and this direct sum is orthogonal to $p_{R \wedge T}^{-1}(0)^\perp$ wrt δ. The fifth and sixth $\Leftrightarrow$'s are elementary properties of geometric orthogonality. □

PROOF OF THEOREM 4.1. Since the correspondence $Q \leftrightarrow p_Q^{-1}(0)$ between $\mathscr{Q}(V)$ and the lattice $\mathscr{L}(V)$ of all subspaces of V [cf. Andersson (1990), Section 4.1] is a lattice anti-isomorphism, it follows that $\mathscr{L} := \{p_Q^{-1}(0) | Q \in \mathscr{Q}\} \subseteq \mathscr{L}(V)$ is a lattice and is anti-isomorphic to $\mathscr{Q}$. If $\sigma \in \mathbf{P}(\mathscr{Q}) \neq \varnothing$, then by Lemma A.3, $\mathscr{L}$ is g.o. wrt $\delta := \sigma^{-1}$. Thus, by Proposition 4.1 of Andersson (1990), $\mathscr{L}$ is distributive; hence so is $\mathscr{Q}$. Conversely, if $\mathscr{Q}$ is distributive, then $\mathbf{P}(\mathscr{Q}) \neq \varnothing$ by Remark 4.1. □

Acknowledgment. We wish to thank a referee and Jesper Madsen for their careful reading of the manuscript and many insightful suggestions.

REFERENCES

ANDERSSON, S. A. (1975). Invariant normal models. *Ann. Statist.* **3** 132–154.

ANDERSSON, S. A. (1990). The lattice structure of orthogonal linear models and orthogonal variance component models. *Scand. J. Statist.* **17** 287–319.

ANDERSSON, S. A., MARDEN, J. I. and PERLMAN, M. D. (1994). Totally ordered multivariate linear models. Unpublished manuscript.

ANDERSSON, S. A. and PERLMAN, M. D. (1991). Lattice-ordered conditional independence models for missing data. *Statist. Probab. Lett.* **12** 465–486.

ANDERSSON, S. A. and PERLMAN, M. D. (1993a). Normal linear models with lattice conditional independence restrictions. In *Proceedings of the International Symposium on Multivariate Analysis and Its Applications* (T. W. Anderson, K.-T. Fang, I. Olkin, eds.). IMS, Hayward, CA.

ANDERSSON, S. A. and PERLMAN, M. D. (1993b). Testing lattice conditional independence models. Unpublished manuscript.

BANERJEE, P. K. and GIRI, N. (1980). On D-, E-, D_A-, and D_M-optimality properties of test procedures of hypotheses concerning the covariance matrix of a normal distribution. In *Multivariate Statistical Analysis* (R. P. Gupta, ed.) 11–19. North-Holland, Amsterdam.

DAS GUPTA, S. (1977). Tests on multiple correlation coefficient and multiple partial correlation coefficient. *J. Multivariate Anal.* **7** 82–88.

DAVEY, B. A. and PRIESTLEY, H. A. (1990). *Introduction to Lattices and Order*. Cambridge Univ. Press.

EATON, M. L. and KARIYA, T. (1983). Multivariate tests with incomplete data. *Ann. Statist.* **11** 654–665.

GIRI, N. (1979). Locally minimax test for multiple correlations. *Canad. J. Statist.* **7** 53–60.

GRÄTZER, G. (1978). *General Lattice Theory*. Birkhäuser, Boston.

KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–52.

LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

MARDEN, J. I. (1981). Invariant tests on covariance matrices. *Ann. Statist.* **9** 1258–1266.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Sample Surveys*. Wiley, New York.

WERMUTH, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.* **75** 963–972.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195