Available online at www.sciencedirect.com
ScienceDirect
Statistics and Probability Letters 78 (2008) 1922–1928
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Conditional independence between two variables given any conditioning subset implies block diagonal covariance matrix for multivariate Gaussian distributions

Guillaume Marrelec*, Habib Benali

Inserm, U678, Paris, F-75013 France
Université Pierre et Marie Curie, Faculté de médecine Pitié-Salpêtrière, Paris, F-75013 France

Received 23 April 2007; received in revised form 3 July 2007; accepted 9 January 2008
Available online 1 February 2008

Abstract

Let $x = (x_{\mathbb{V}})$ be a multivariate Gaussian variable with covariance matrix Σ. For i and j in $\mathbb{V}$, we show that if the conditional covariance between x_i and x_j given any conditioning set $\mathbb{K} \subset \mathbb{V} \setminus \{i, j\}$ is equal to zero, then Σ is block diagonal and i and j belong to two different blocks.

1. Introduction

As pointed out by Dawid (1998), the concept of conditional independence is believed to be the fundamental knowledge in the process of scientific inference. For multivariate Gaussian variables, conditional independence is quantified by conditional covariance. Investigation of such coefficients have led to a better characterization of interactions between variables, in particular through the use of conditional independence graphs (Whittaker, 1990; Lauritzen, 1996; Edwards, 2000). Marginal correlation coefficients have also been examined through covariance graphs (Kauermann, 1996; Edwards, 2000). It would be interesting to generalize these approaches by simultaneously considering *all possible* conditional covariances for a given pair of variables. For instance, consider the case of a three-dimensional Gaussian variable $x = (x_1, x_2, x_3)$ with covariance matrix Σ. If $\mathrm{Corr}[x_1, x_2|x_3] = 0$, then $\mathrm{Corr}[x_1, x_2] = \mathrm{Corr}[x_1, x_3] \cdot \mathrm{Corr}[x_2, x_3]$ (Wermuth, 1976; Whittaker, 1990, Proposition 2.6.5). If one furthermore has $\mathrm{Corr}[x_1, x_2] = 0$, it directly comes out that either $\mathrm{Corr}[x_1, x_3] = 0$ or $\mathrm{Corr}[x_2, x_3] = 0$. In other words, the following yields:

$$\{\mathrm{Corr}[x_1, x_2] = 0 \text{ and } \mathrm{Corr}[x_1, x_2|x_3] = 0\} \Rightarrow \ \Sigma \text{ is block diagonal.}$$

To our knowledge, no generalization of such a result has been shown yet. This paper is a first step in this direction. We prove a result that demonstrates how this approach can inform us regarding the global pattern of interaction and shed light into the structure of the variables.

* Corresponding address: CHU Pitié-Salpêtrière, 91 boulevard de l'Hôpital, 75634 Paris Cedex 13, France.
E-mail address: marrelec@imed.jussieu.fr (G. Marrelec).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2008.01.060

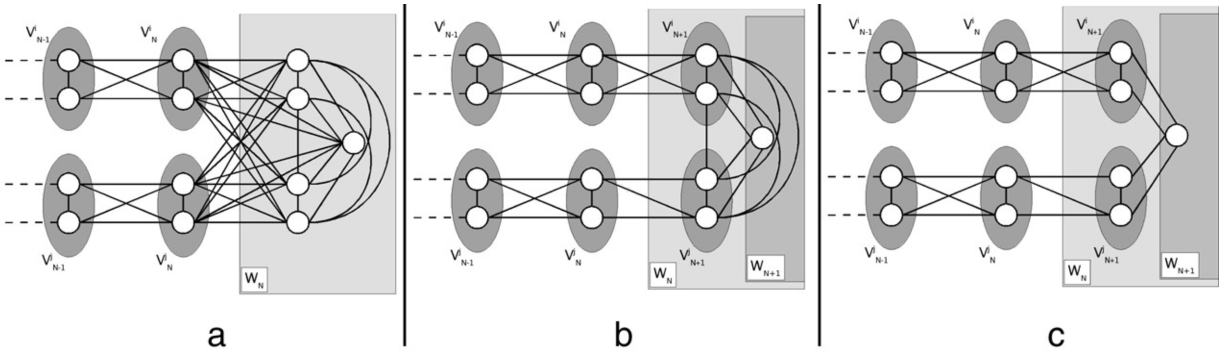

Fig. 1. Sketch of proof. From an original partitioning of $\mathbb{V}$ into $\{\mathbb{V}_0^i, \ldots, \mathbb{V}_N^i, \mathbb{V}_0^j, \ldots, \mathbb{V}_N^j, \mathbb{W}_N\}$ (a), we proceed as follows. We first show that for all elements of $\mathbb{W}_N$ there can be no marginal covariance with both $\mathbb{V}_N^i$ and $\mathbb{V}_N^j$. We then partition $\mathbb{W}_N$ into elements that covariate with $\mathbb{V}_N^i$ (gathered in $\mathbb{V}_{N+1}^i$), elements that covariate with $\mathbb{V}_N^j$ (gathered in $\mathbb{V}_{N+1}^j$), and elements that covariate with neither (gathered in $\mathbb{W}_{N+1}$) (b). Last, we show that elements of $\mathbb{V}_{N+1}^i$ and $\mathbb{V}_{N+1}^j$ must have zero marginal covariance (c).

2. Main theorem

Let $\mathbb{V}$ be a finite set and $x = x_{\mathbb{V}} = (x_v)_{v \in \mathbb{V}}$ be a multivariate Gaussian variable indexed on $\mathbb{V}$ with covariance matrix Σ.

Theorem 1. *Let i and j be two elements of $\mathbb{V}$ and further assume that x_i and x_j are conditionally independent given any set of remaining variables, i.e.,*

$$\forall\, \mathbb{K} \subset \mathbb{V} \setminus \{i, j\} \quad \mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}] = 0. \tag{1}$$

Then Σ is block diagonal and i and j belong to two different blocks.

Sole consideration of marginal and/or partial covariance is not sufficient to provide this result, for there exist covariance matrices that are not block diagonal while including variables for which $\mathrm{Cov}[x_i, x_j] = 0$ and/or $\mathrm{Cov}[x_i, x_j | x_{\mathbb{V}\setminus\{1,j\}}] = 0$.

This result can be established by successive examination of conditional independence constraints (see Fig. 1 for a graphical sketch of proof). First, $\mathrm{Corr}[x_i, x_j] = 0$ and, hence, $\Sigma_{i,j} = 0$. We also have $\mathrm{Cov}[x_i, x_j | x_k] = 0$ for any $k \in \mathbb{V} \setminus \{i, j\}$. Since $\Sigma_{i,j} = 0$, the general formula for conditional covariance (Anderson, 1958, p. 28) simplifies to

$$\mathrm{Cov}[x_i, x_j | x_k] = -\frac{\Sigma_{i,k}\Sigma_{k,j}}{\Sigma_{k,k}}.$$

For $\mathrm{Cov}[x_i, x_j | x_k]$ to be equal to zero, we must then have $\Sigma_{i,k}\Sigma_{k,j} = 0$, i.e., either $\Sigma_{i,k} = 0$ or $\Sigma_{j,k} = 0$. This line of reasoning being valid for any $k \notin \{i, j\}$, it is possible to separate $\mathbb{V} \setminus \{i, j\}$ into three sets: $\mathbb{V}_1^i$ such that $\Sigma_{i,k} \neq 0$ and $\Sigma_{j,k} = 0$ for $k \in \mathbb{V}_1^i$; $\mathbb{V}_1^j$ such that $\Sigma_{i,k} = 0$ and $\Sigma_{j,k} \neq 0$ for $k \in \mathbb{V}_1^j$; and $\mathbb{W}_1$ such that $\Sigma_{i,k} = 0$ and $\Sigma_{j,k} = 0$ for $k \in \mathbb{W}_1$. Let then be $\mathbb{K} = \{k, l\}$ with $k \in \mathbb{V}_1^i$ and $l \in \mathbb{V}_1^j$. $\mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}]$ is given by (see Eq. (A.1))

$$\mathrm{Cov}[x_i, x_j \mid x_{\mathbb{K}}] = -\sum_{a,b \in \mathbb{K}} \left(\Sigma_{i\cup j,\mathbb{K}}\right)_{i,a} \frac{(-1)^{\mathrm{pos}_{\mathbb{K}}(a)+\mathrm{pos}_{\mathbb{K}}(b)} \det\left[\Sigma_{\mathbb{K}\setminus\{b\},\mathbb{K}\setminus\{a\}}\right]}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]} \left(\Sigma_{\mathbb{K},i\cup j}\right)_{b,j},$$

where $\mathrm{pos}_{\mathbb{K}}(a)$ stands for the position of a in $\mathbb{K}$. Since $k \in \mathbb{V}_1^i$ and $l \in \mathbb{V}_1^j$, we have $\Sigma_{i,l} = \Sigma_{j,k} = 0$ and $\mathrm{Cov}[x_i, x_j \mid x_{\mathbb{K}}]$ boils down to

$$\mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}] = \frac{\Sigma_{i,k}\Sigma_{k,l}\Sigma_{l,j}}{\Sigma_{k,k}\Sigma_{l,l} - \Sigma_{k,k}^2}.$$

Since we must also have $\mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}] = 0$ according to our hypothesis, this equation leads to $\Sigma_{k,l} = 0$, given that $\Sigma_{i,k}$ and $\Sigma_{j,l}$ are different from zero. Elements of $\mathbb{V}_1^i$ (resp. $\mathbb{V}_1^j$) have hence a zero marginal correlation to both j (resp. i) and all elements of $\mathbb{V}_1^j$ (resp. $\mathbb{V}_1^i$).

We then proceed by induction. Assume that there exist $2(N+1)$ subsets $\mathbb{V}_n^i$ and $\mathbb{V}_n^j$, with $n = 0, \ldots, N$, and one set $\mathbb{W}_N$ of $\mathbb{V}$ such that

- $\mathbb{V}_0^i = \{i\}$ and $\mathbb{V}_0^j = \{j\}$;
- $\{\mathbb{V}_0^i, \ldots, \mathbb{V}_N^i, \mathbb{V}_0^j, \ldots, \mathbb{V}_N^j, \mathbb{W}_N\}$ is a partition of $\mathbb{V}$.
- nonzero marginal correlations can only be found between $\mathbb{V}_{n-1}^i$ and $\mathbb{V}_n^i$, between $\mathbb{V}_{n-1}^j$ and $\mathbb{V}_n^j$, or between $\mathbb{W}_N$ and $\{\mathbb{V}_N^i, \mathbb{V}_N^j\}$.
- all marginal correlations between $\mathbb{V}_{n-1}^i$ and $\mathbb{V}_n^i$, as well as between $\mathbb{V}_{n-1}^j$ and $\mathbb{V}_n^j$ are different from zero.

Since we proved that $\Sigma_{i,j} = 0$, $\Sigma_{i,l} = 0$ for $l \in \mathbb{V}_1^j$, $\Sigma_{j,k} = 0$ for $k \in \mathbb{V}_1^i$, $\Sigma_{k,l} = 0$ for $(k,l) \in \mathbb{V}_1^j \times \mathbb{V}_1^j$ and constructed $\mathbb{V}_1^i$ and $\mathbb{V}_1^j$ so that $\Sigma_{i,k} \neq 0$ for $k \in \mathbb{V}_1^i$ and $\Sigma_{j,l} \neq 0$ for $l \in \mathbb{V}_1^j$, the assumption holds for $N = 1$. We now assume that it also holds for a given $N \geq 1$. If $\mathbb{W}_N$ is empty, then the process stops. Otherwise, the first step consists of setting $\mathbb{K} = \{k_1, l_1, \ldots, k_N, l_N, m\}$, with $(k_n, l_n) \in \mathbb{V}_n^i \times \mathbb{V}_n^j$ for $n = 1 \ldots, N$, and $m \in \mathbb{W}_N$. Given the assumption of independence between x_i and x_j, we must have $\mathrm{Cov}[x_i, x_j | x_\mathbb{K}] = 0$. This conditional covariance coefficient is equal to (cf. Eq. (A.2))

$$\frac{\Sigma_{i,k_1}\Sigma_{j,l_1}\left[\prod_{n=1,\ldots,N-1}\Sigma_{k_n,k_{n+1}}\Sigma_{l_n,l_{n+1}}\right]\Sigma_{k_N,m}\Sigma_{l_N,m}}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]}$$

and is equal to zero if and only if $\Sigma_{k_N,m}\Sigma_{l_N,m} = 0$, since, by construction all $\Sigma_{k_n,k_{n+1}}$ and $\Sigma_{l_n,l_{n+1}}$ are different from zero. It is then possible to separate $\mathbb{W}_N$ into three sets: $\mathbb{V}_{N+1}^i$ such that $\Sigma_{k_N,m} \neq 0$ and $\Sigma_{l_N,m} = 0$ for all $m \in \mathbb{V}_{N+1}^i$; $\mathbb{V}_{N+1}^j$ such that $\Sigma_{k_N,m} = 0$ and $\Sigma_{l_N,m} \neq 0$ for all $m \in \mathbb{V}_{N+1}^j$; and $\mathbb{W}_{N+1}$ such that $\Sigma_{k_N,m} = \Sigma_{l_N,m} = 0$ for all $m \in \mathbb{W}_{N+1}$. It now remains to prove that we have $\Sigma_{k,l} = 0$ for $(k,l) \in \mathbb{V}_{N+1}^i \times \mathbb{V}_{N+1}^j$. To this aim, set $\mathbb{K} = \{k_1, l_1, \ldots, k_{N+1}, l_{N+1}\}$ with $(k_n, l_n) \in \mathbb{V}_n^i \times \mathbb{V}_n^j$ for $n = 1 \ldots, N+1$. Since x_i and x_j are independent, we must have $\mathrm{Cov}[x_i, x_j | x_\mathbb{K}] = 0$. This quantity being equal to (see Eq. (A.3))

$$\mathrm{Cov}\left[x_i, x_j | x_\mathbb{K}\right] = \frac{\Sigma_{i,k_1}\Sigma_{j,l_1}\left[\prod_{n=1,\ldots,N}\Sigma_{k_n,k_{n+1}}\Sigma_{l_n,l_{n+1}}\right]\Sigma_{k_{N+1},l_{N+1}}}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]},$$

it is equal to zero if and only if $\Sigma_{k_{N+1},l_{N+1}} = 0$. The assumption is therefore also valid for $N+1$.

The sequence $(\mathbb{W}_N)$ is of decreasing cardinal. $\mathbb{V}$ being a finite set, there exists a step N_0 for which $\mathbb{W}_{N_0}$ is empty: the process ends there. Set $\mathbb{V}^i = \{V_0^i, \ldots, \mathbb{V}_{N_0}^i\}$ and $\mathbb{V}^j = \{\mathbb{V}_0^j, \ldots, \mathbb{V}_{N_0}^j\}$. $\{\mathbb{V}^i, \mathbb{V}^j\}$ is hence a partition of $\mathbb{V}$ for which there exists no marginal correlation between an element of $\mathbb{V}^i$ and an element of $\mathbb{V}^j$. Consequently, the covariance matrix of x has the following structural form:

$$\begin{pmatrix} \Sigma_{\mathbb{V}^i,\mathbb{V}^i} & 0 \\ 0 & \Sigma_{\mathbb{V}^j,\mathbb{V}^j} \end{pmatrix},$$

thereby proving the theorem.

3. Discussion and perspectives

In this paper, we considered $x = x_\mathbb{V} = (x_v)_{v\in\mathbb{V}}$ a multivariate Gaussian variable with covariance matrix Σ. For i and j in $\mathbb{V}$, we showed that if the conditional covariance between x_i and x_j given any conditioning set $\mathbb{K} \subset \mathbb{V} \setminus \{i, j\}$ is equal to zero, then Σ is block diagonal and i and j belonged to two different blocks. Note that the converse of this theorem is straightforward. Indeed, if one considers that the covariance matrix Σ is block diagonal, then any conditional covariance between variables belonging to two different blocks is equal to zero according to Eq. (A.1).

Theorem 1 shows that, for multivariate Gaussian variables, there is a clear separation between two variables x_i and x_j that are independent with regard to any conditioning subset, and that this separation also applies to all other variables, which are either "with" x_i or "with" x_j. Consequently, their effect can be analyzed independently in one block of variables or the other.

Interestingly, this result nicely relates two distinct properties of Gaussian distributions. The block diagonal property of the covariance matrix is clearly a global feature of Gaussian probability distributions. By contrast, the relationship of "complete independence" (i.e., conditioned on all subsets) is rather a local description and characterization of the interaction structure between variables, since the definition gives a particular role to x_i and x_j. This perspective differs from the common approach, where one usually sets a "level" of conditioning (marginal for covariance graphs, partial for conditional independence graphs) and then varies the two variables on which correlation is calculated. In this "dual" approach, the definition does not so much depend on the conditioning set than on the variables whose conditional covariance we examine. We mainly focus on the independence pattern that can be exhibited with a single pair of variables and its potential implications onto the global structure. We believe that there is much to gain by analyzing variables from this perspective and hope to be able to provide further results along the same lines in the near future.

Acknowledgment

We are in debt to an anonymous referee for pointing out that the result exposed here is well known for three-dimensional Gaussian variables.

Appendix. Calculation of $\mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}]$

The conditional covariance between i and j given $\mathbb{K}$ reads (Anderson, 1958, p. 28)

$$\mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}] = \Sigma_{i,j} - \sum_{a,b\in\mathbb{K}} \left(\Sigma_{i\cup j,\mathbb{K}}\right)_{i,a} \left[\left(\Sigma_{\mathbb{K},\mathbb{K}}\right)^{-1}\right]_{a,b} \left(\Sigma_{\mathbb{K},i\cup j}\right)_{b,j}.$$

Calculating $\left(\Sigma_{\mathbb{K},\mathbb{K}}\right)^{-1}$ from the adjoint matrix (Horn and Johnson, 1999) yields for $\mathrm{Cov}[x_i, x_j | x_{\mathbb{K}}]$:

$$cov[x_i, x_j \mid x_{\mathbb{K}}] = \Sigma_{i,j} - \sum_{a,b\in\mathbb{K}} \left(\Sigma_{i\cup j,\mathbb{K}}\right)_{i,a} \frac{(-1)^{\mathrm{pos}_{\mathbb{K}}(a)+\mathrm{pos}_{\mathbb{K}}(b)} \det\left[\Sigma_{\mathbb{K}\setminus\{b\},\mathbb{K}\setminus\{a\}}\right]}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]} \left(\Sigma_{\mathbb{K},i\cup j}\right)_{b,j}, \tag{A.1}$$

where $\mathrm{pos}_{\mathbb{K}}(a)$ stands for the position of a in $\mathbb{K}$. From now on, we also assume that there exist $2(N+1)+1$ subsets of $\mathbb{V}$, namely $\mathbb{V}_n^i$, $\mathbb{V}_n^j$, with $n = 0, \dots, N$, and $\mathbb{W}_N$, respecting the conditions detailed on page 3.

First, for $N \geq 1$, set $\mathbb{K} = \{k_1, l_1, \dots, k_N, l_N, m\}$, $k_n \in \mathbb{V}_n^i$ and $l_n \in \mathbb{V}_n^j$ for $n = 1\dots, N$, and $m \in \mathbb{W}_N$. By construction, only elements in $\mathbb{V}_1^i$ (resp. $\mathbb{V}_1^j$) have nonzero marginal covariance with i (resp. j). Consequently, the sum in Eq. (A.1) can be simplified into

$$\Sigma_{i,k_1} \frac{(-1)^{\mathrm{pos}_{\mathbb{K}}(a)+\mathrm{pos}_{\mathbb{K}}(b)} \det\left[\Sigma_{\mathbb{K}\setminus\{l_1\},\mathbb{K}\setminus\{k_1\}}\right]}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]} \Sigma_{j,l_1}.$$

Given the definition of $\mathbb{K}$, $\Sigma_{\mathbb{K},\mathbb{K}}$, $\Sigma_{\mathbb{K}\setminus\{l_1\},\mathbb{K}\setminus\{k_1\}}$, and the determinant of the latter matrix respectively read

$$\Sigma_{\mathbb{K},\mathbb{K}} = \begin{pmatrix} \Sigma_{k_1,k_1} & 0 & \Sigma_{k_1,k_2} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_{l_1,l_1} & 0 & \Sigma_{l_1,l_2} & 0 & 0 & 0 & & \\ \Sigma_{k_1,k_2} & 0 & \Sigma_{k_2,k_2} & 0 & \Sigma_{k_2,k_3} & 0 & 0 & & \vdots \\ 0 & \Sigma_{l_1,l_2} & 0 & \Sigma_{l_2,l_2} & 0 & \Sigma_{l_2,l_3} & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 \\ \vdots & & 0 & 0 & \Sigma_{k_{N-1},k_N} & 0 & \Sigma_{k_N,k_N} & 0 & \Sigma_{k_N,m} \\ & & 0 & 0 & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_N,l_N} & \Sigma_{l_N,m} \\ 0 & \cdots & 0 & 0 & 0 & 0 & \Sigma_{k_N,m} & \Sigma_{l_N,m} & \Sigma_{m,m} \end{pmatrix},$$

$$\Sigma_{\mathbb{K}\backslash\{l_1\},\mathbb{K}\backslash\{k_1\}} = \begin{pmatrix} 0 & \Sigma_{k_1,k_2} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_{k_2,k_2} & 0 & \Sigma_{k_2,k_3} & 0 & 0 & 0 & & \\ \Sigma_{l_1,l_2} & 0 & \Sigma_{l_2,l_2} & 0 & \Sigma_{l_2,l_3} & 0 & 0 & & \vdots \\ 0 & \Sigma_{k_2,k_3} & 0 & \Sigma_{k_3,k_3} & 0 & \Sigma_{k_3,k_4} & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 \\ \vdots & & 0 & 0 & \Sigma_{k_{N-1},k_N} & 0 & \Sigma_{k_N,k_N} & 0 & \Sigma_{k_N,m} \\ & & 0 & 0 & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_N,l_N} & \Sigma_{l_N,m} \\ 0 & \cdots & 0 & 0 & 0 & 0 & \Sigma_{k_N,m} & \Sigma_{l_N,m} & \Sigma_{m,m} \end{pmatrix},$$

and

$$\det\left[\Sigma_{\mathbb{K}\backslash\{l_1\},\mathbb{K}\backslash\{k_1\}}\right] = -\Sigma_{k_1,k_2} \det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1\},\mathbb{K}\backslash\{k_1,k_2\}}\right).$$

We also have

$$\Sigma_{\mathbb{K}\backslash\{l_1,k_1\},\mathbb{K}\backslash\{k_1,k_2\}} = \begin{pmatrix} 0 & 0 & \Sigma_{k_2,k_3} & 0 & 0 & 0 & 0 & \cdots & 0 \\ \Sigma_{l_1,l_2} & \Sigma_{l_2,l_2} & 0 & \Sigma_{l_2,l_3} & 0 & 0 & 0 & & \\ 0 & 0 & \Sigma_{k_3,k_3} & 0 & \Sigma_{k_3,k_4} & 0 & 0 & & \vdots \\ 0 & \Sigma_{l_2,l_3} & 0 & \Sigma_{l_3,l_3} & 0 & \Sigma_{l_3,l_4} & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 \\ \vdots & & 0 & 0 & \Sigma_{k_{N-1},k_N} & 0 & \Sigma_{k_N,k_N} & 0 & \Sigma_{k_N,m} \\ & & 0 & 0 & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_N,l_N} & \Sigma_{l_N,m} \\ 0 & \cdots & 0 & 0 & 0 & 0 & \Sigma_{k_N,m} & \Sigma_{l_N,m} & \Sigma_{m,m} \end{pmatrix},$$

and, hence,

$$\det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1\},\mathbb{K}\backslash\{k_1,k_2\}}\right) = -\Sigma_{l_1,l_2} \det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,l_2\},\mathbb{K}\backslash\{k_1,l_1,k_2\}}\right).$$

This leads to

$$\det\left[\Sigma_{\mathbb{K}\backslash\{l_1\},\mathbb{K}\backslash\{k_1\}}\right] = \Sigma_{k_1,k_2}\Sigma_{l_1,l_2} \det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,l_2\},\mathbb{K}\backslash\{k_1,l_1,k_2\}}\right),$$

with

$$\Sigma_{\mathbb{K}\backslash\{l_1,k_1,l_2\},\mathbb{K}\backslash\{k_1,k_2,l_1\}} = \begin{pmatrix} 0 & \Sigma_{k_2,k_3} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_{k_3,k_3} & 0 & \Sigma_{k_3,k_4} & 0 & 0 & 0 & & \\ \Sigma_{l_2,l_3} & 0 & \Sigma_{l_3,l_3} & 0 & \Sigma_{l_3,l_4} & 0 & 0 & & \vdots \\ 0 & \Sigma_{l_3,l_4} & 0 & \Sigma_{k_4,k_4} & 0 & \Sigma_{k_4,k_5} & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 & \Sigma_{l_{N-1},l_{N-1}} & 0 \\ \vdots & & 0 & 0 & \Sigma_{k_{N-1},k_N} & 0 & \Sigma_{k_N,k_N} & 0 & \Sigma_{k_N,m} \\ & & 0 & 0 & 0 & \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_N,l_N} & \Sigma_{l_N,m} \\ 0 & \cdots & 0 & 0 & 0 & 0 & \Sigma_{k_N,m} & \Sigma_{l_N,m} & \Sigma_{m,m} \end{pmatrix},$$

which is of the same form as $\Sigma_{\mathbb{K}\backslash\{l_1\},\mathbb{K}\backslash\{k_1\}}$. Consequently, a similar calculation shows that

$$\det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,l_2\},\mathbb{K}\backslash\{k_1,l_1,k_2\}}\right) = \Sigma_{k_2,k_3}\Sigma_{l_2,l_3} \det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,l_2,k_2,l_3\},\mathbb{K}\backslash\{k_1,l_1,k_2,l_2,k_3\}}\right),$$

and, by induction, one can hence easily show that, for all $n \geq 2$,

$$\det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,\ldots,l_{n-1},k_{n-1},l_n\},\mathbb{K}\backslash\{k_1,l_1,\ldots,k_{n-1},l_{n-1},k_n\}}\right)$$
$$= \Sigma_{k_n,k_{n+1}}\Sigma_{l_n,l_{n+1}}\det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,\ldots,l_n,k_n,l_{n+1}\},\mathbb{K}\backslash\{k_1,l_1,\ldots,k_n,l_n,k_{n+1},l_n\}}\right).$$

We hence obtain that

$$\det\left[\Sigma_{\mathbb{K}\backslash\{l_1\},\mathbb{K}\backslash\{k_1\}}\right] = \det\left(\Sigma_{\mathbb{K}\backslash\{l_1,k_1,\ldots,k_{N-2},l_{N-2},l_{N-1}\},\mathbb{K}\backslash\{k_1,l_1,\ldots,k_{N-2},l_{N-2},k_{N-1}\}}\right)\prod_{n=1,\ldots,N-2}\Sigma_{k_n,k_{n+1}}\Sigma_{l_n,l_{n+1}},$$

where the matrix whose determinant is calculated is equal to

$$\begin{aligned}\Sigma_{\mathbb{K}\backslash\{l_1,k_1,\ldots,k_{N-2},l_{N-2},l_{N-1}\},\mathbb{K}\backslash\{k_1,l_1,\ldots,k_{N-2},l_{N-2},k_{N-1}\}} &= \Sigma_{\{k_{N-1},k_N,l_n,m\},\{l_{N-1},k_N,l_N,m\}}\\ &= \begin{pmatrix} 0 & \Sigma_{k_{N-1},k_N} & 0 & 0\\ 0 & \Sigma_{k_N,k_N} & 0 & \Sigma_{k_N,m}\\ \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_N,l_N} & \Sigma_{l_N,m}\\ 0 & \Sigma_{k_N,m} & \Sigma_{l_N,m} & \Sigma_{m,m}\end{pmatrix}.\end{aligned}$$

The determinant of this matrix can be obtained by a similar argument as previously developed:

$$\begin{aligned}\det\left(\Sigma_{\{k_{N-1},k_N,l_n,m\},\{l_{N-1},k_N,l_N,m\}}\right) &= -\Sigma_{k_{N-1},k_N}\begin{vmatrix} 0 & 0 & \Sigma_{k_N,m}\\ \Sigma_{l_{N-1},l_N} & \Sigma_{l_N,l_N} & \Sigma_{l_N,m}\\ 0 & \Sigma_{l_N,m} & \Sigma_{m,m}\end{vmatrix}\\ &= \Sigma_{k_{N-1},k_N}\Sigma_{l_{N-1},l_N}\begin{vmatrix} 0 & \Sigma_{k_N,m}\\ \Sigma_{l_N,m} & \Sigma_{m,m}\end{vmatrix}\\ &= -\Sigma_{k_{N-1},k_N}\Sigma_{l_{N-1},l_N}\Sigma_{k_N,m}\Sigma_{l_N,m}.\end{aligned}$$

We finally have

$$\det\left[\Sigma_{\mathbb{K}\backslash\{l_1\},\mathbb{K}\backslash\{k_1\}}\right] = \Sigma_{k_N,m}\Sigma_{l_N,m}\prod_{n=1,\ldots,N-1}\Sigma_{k_n,k_{n+1}}\Sigma_{l_n,l_{n+1}},$$

and, in conclusion, for $\mathbb{K} = \{k_1, l_1, \ldots, k_N, l_N, m\}$, we obtain for $\mathrm{Cov}\left[x_i, x_j|x_{\mathbb{K}}\right]$

$$\frac{\Sigma_{i,k_1}\left[\prod_{n=1,\ldots,N-1}\Sigma_{k_n,k_{n+1}}\right]\Sigma_{k_N,m}\cdot\Sigma_{j,l_1}\left[\prod_{n=1,\ldots,N-1}\Sigma_{l_n,l_{n+1}}\right]\Sigma_{l_N,m}}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]}. \tag{A.2}$$

The second case is rather similar to the first one. Set $\mathbb{K} = \{k_1, l_1, \ldots, k_{N+1}, l_{N+1}\}$, with $N \geq 1$, $k_n \in \mathbb{V}_n^i$ and $l_n \in \mathbb{V}_n^j$ for $n = 1\ldots, N+1$. The previous line of reasoning can be applied in this case too, except that $\Sigma_{\mathbb{K}\backslash\{l_1,k_1,\ldots,l_{N-1}\},\mathbb{K}\backslash\{k_1,l_1,\ldots,k_{N-1}\}}$ reads

$$\Sigma_{\{k_{N-1},k_N,l_N,k_{N+1},l_{N+1}\},\{l_{N-1},k_N,l_N,k_{N+1},l_{N+1}\}} = \begin{pmatrix} 0 & \Sigma_{k_{N-1},k_N} & 0 & 0 & 0\\ 0 & \Sigma_{k_N,k_N} & 0 & \Sigma_{k_N,k_{N+1}} & 0\\ \Sigma_{l_{N-1},l_N} & 0 & \Sigma_{l_N,l_N} & 0 & \Sigma_{l_N,l_{N+1}}\\ 0 & \Sigma_{k_N,k_{N+1}} & 0 & \Sigma_{k_{N+1},k_{N+1}} & \Sigma_{k_{N+1},l_{N+1}}\\ 0 & 0 & \Sigma_{l_N,l_{N+1}} & \Sigma_{k_{N+1},l_{N+1}} & \Sigma_{l_N,l_N}\end{pmatrix},$$

leading to a determinant of $\Sigma_{\mathbb{K}\backslash\{l_1,k_1,\ldots,,l_{N-1}\},\mathbb{K}\backslash\{k_1,l_1,\ldots,k_{N-1}\}}$ equal to

$$\begin{aligned}&= -\Sigma_{k_{N-1},k_N}\begin{vmatrix} 0 & 0 & \Sigma_{k_N,k_{N+1}} & 0\\ \Sigma_{l_{N-1},l_N} & \Sigma_{l_N,l_N} & 0 & \Sigma_{l_N,l_{N+1}}\\ 0 & 0 & \Sigma_{k_{N+1},k_{N+1}} & \Sigma_{k_{N+1},l_{N+1}}\\ 0 & \Sigma_{l_N,l_{N+1}} & 0 & \Sigma_{l_N,l_N}\end{vmatrix}\\ &= \Sigma_{k_{N-1},k_N}\Sigma_{l_{N-1},l_N}\begin{vmatrix} 0 & \Sigma_{k_N,k_{N+1}} & 0\\ 0 & \Sigma_{k_{N+1},k_{N+1}} & \Sigma_{k_{N+1},l_{N+1}}\\ \Sigma_{l_N,l_{N+1}} & 0 & \Sigma_{l_N,l_N}\end{vmatrix}\end{aligned}$$

$$= -\Sigma_{k_{N-1},k_N}\,\Sigma_{l_{N-1},l_N}\,\Sigma_{k_N,k_{N+1}} \begin{vmatrix} 0 & \Sigma_{k_{N+1},l_{N+1}} \\ \Sigma_{l_N,l_{N+1}} & \Sigma_{l_N,l_N} \end{vmatrix}$$
$$= \Sigma_{k_{N-1},k_N}\,\Sigma_{l_{N-1},l_N}\,\Sigma_{k_N,k_{N+1}}\,\Sigma_{l_N,l_{N+1}}\,\Sigma_{k_{N+1},l_{N+1}}.$$

Finally, $\mathrm{Cov}\left[x_i, x_j | x_{\mathbb{K}}\right]$ reads

$$\frac{\Sigma_{i,k_1}\left[\prod_{n=1,\dots,N}\Sigma_{k_n,k_{n+1}}\right]\cdot\Sigma_{j,l_1}\left[\prod_{n=1,\dots,N}\Sigma_{l_n,l_{n+1}}\right]\cdot\Sigma_{k_{N+1},l_{N+1}}}{\det\left[\Sigma_{\mathbb{K},\mathbb{K}}\right]}. \tag{A.3}$$

References

Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. Wiley Publications in Statistics. John Wiley and Sons, New York.
Dawid, A.P., 1998. Conditional independence. In: Kotz, S., Read, C.B., Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, vol. 2. Wiley, pp. 146–155.
Edwards, D., 2000. Introduction to Graphical Modelling, 2nd ed. In: Springer Texts in Statistics. Springer, New York.
Horn, R.A., Johnson, C.R., 1999. Matrix Analysis. Cambridge University Press.
Kauermann, G., 1996. On a dualization of graphical Gaussian models. Scand. J. Stat. 23, 105–116.
Lauritzen, S.L., 1996. Graphical Models. Oxford University Press, Oxford.
Wermuth, N., 1976. Analogies between multiplicative models in contigency tables and covariance selection. Biometrics 32, 95–108.
Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics. J. Wiley and Sons, Chichester.