Bernoulli **23**(1), 2017, 626–644
DOI: 10.3150/15-BEJ759

Algebraic representations of Gaussian Markov combinations

M. SOFIA MASSA[1] and EVA RICCOMAGNO[2]

[1]*Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, United Kingdom, OX1 3TG. E-mail: massa@stats.ox.ac.uk*
[2]*Dipartimento di Matematica, Università degli Studi di Genova, Via Dodecaneso, 35, 16146 Genova, Italy. E-mail: riccomagno@dima.unige.it*

Markov combinations for structural meta-analysis problems provide a way of constructing a statistical model that takes into account two or more marginal distributions by imposing conditional independence constraints between the variables that are not jointly observed. This paper considers Gaussian distributions and discusses how the covariance and concentration matrices of the different combinations can be found via matrix operations. In essence, all these Markov combinations correspond to finding a positive definite completion of the covariance matrix over the set of random variables of interest and respecting the constraints imposed by each Markov combination. The paper further shows the potential of investigating the properties of the combinations via algebraic statistics tools. An illustrative application will motivate the importance of solving problems of this type.

Keywords: algebraic statistics; conditional independence; Gaussian graphical models; Markov combinations

1. Introduction

Markov combinations of distributions (or families of distributions) are operators that define a new distribution (or family) starting from two initial ones by imposing conditional independences between the variables that are not jointly observed. If the original families are Gaussian, the Markov combinations are Gaussian themselves and if no other constraints are added, then the only constraints are the conditional independencies arising from imposing the Markov combination. If the initial families encode other types of constraints between their variables, then those constraints need to be taken into account as well. For example, if the families are Gaussian graphical models, a Markov combination takes into account all constraints imposed by each Gaussian graphical model and proposes a new model that combines the information provided by both models. Different Markov combinations will make different use of the covariance/concentration structures of the models to be combined.

The Markov combination for consistent distributions and the meta-Markov combination for meta-consistent families of distributions were introduced by [3] while defining a framework for Bayesian inference for graphical models. Given the set of cliques and separators of a decomposable graph and a set of pairwise consistent distributions defined over the cliques of the graph, the Markov combination can be used to construct a Markov distribution with respect to the graph.

In [10], the setting of Markov combinations was exploited for building distributions over a union of two sets of random variables specifying two distinct statistical models. The lower, upper, super Markov combinations were introduced together with an investigation of their use when the initial information is represented by Gaussian graphical models. In this setting, the combination of evidence was named structural meta-analysis because the initial information was represented by structures of conditional independence. In [9], it was shown via simulation studies that combination of information may be effective also in model selections problems.

The aim of this paper is to provide an algebraic characterization of these types of combinations for Gaussian distributions and families of distribution. We will present a matrix representation of the Markov combination for distribution and the lower, upper and super Markov combinations for families of distributions, showing how the covariance and concentration matrices of the different combinations can be computed from the covariance and concentration matrices of the initial Gaussian models. These types of combinations correspond to finding a positive definite completion of suitable defined covariance matrices over the set of random variables of interest. Further an algebraic representation of each combination will be given in terms of polynomial ideals [11]. This is particularly useful when we want to look more in depth at the combination of equality and inequality constraints imposed by conditional independence relations, stationary constraints or positive definiteness requirements, for example. Finally, an illustrative example using real data will be presented.

2. Preliminaries

This section reviews some background notation and definitions. It briefly revisits some types of Markov combinations as defined in [10] and some aspects of conditional independence.

2.1. Markov combinations

Assume $v \in [n] = \{1, \dots, n\}$ and let Y_v denote the random variable indexed by v. For $A, B \subseteq [n]$, let Y_A denote the vector $(Y_v)_{v \in A}$. If f is a density over $Y_{[n]}$, let f_A be the marginal density for Y_A and $f_{A|B}$ the conditional density of $Y_{A \setminus B}$ given $Y_B = y_B$ where $y_B = (y_b)_{b \in B}$.

A non-degenerate multivariate Gaussian distribution $\mathcal{N}_n(\mu, \Sigma)$ is specified by its mean vector μ and its covariance matrix $\Sigma \in \mathcal{S}_n^+$, where $\mathcal{S}_n^+$ is the set of symmetric positive definite matrices with n rows. Assume two non-degenerate Gaussian families $\mathcal{F}$ and $\mathcal{G}$ for the random vectors $Y_{A \cup C}$ and $Y_{C \cup B}$, respectively with A, B, C a partition of $[n]$. Markov combinations are defined for the joint random vector $Y_{A \cup C \cup B}$. If $\mathcal{F}$ and $\mathcal{G}$ are Gaussian, so are their combinations and this is the setting of the paper.

Two distributions $f \in \mathcal{F}$ and $g \in \mathcal{G}$ are said to be consistent if $f_C = g_C$. The Markov combination of two consistent densities f and g is defined as $f \star g = f \cdot g / f_C$. See Definitions 3.1 and 4.1 in [10] and references therein. Let $\mathcal{F}^{\downarrow A}$ denote the marginal family of distributions induced by $\mathcal{F}$ on the vector $(Y_a)_{a \in A}$. The families of distributions $\mathcal{F}$ and $\mathcal{G}$ are said to be meta-consistent if $\mathcal{F}^{\downarrow C} = \mathcal{G}^{\downarrow C}$.

Definition 2.1. *Let A, B, C be a partition of $[n]$ and let $\mathcal{F}$ and $\mathcal{G}$ be two Gaussian families, defined for the random vectors $Y_{A \cup C}$ and $Y_{C \cup B}$, respectively. Their Markov combinations are defined for $Y_{A \cup C \cup B}$ as follows.*

Meta ([3]). *If $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent, their meta-Markov combination is defined as*

$$\mathcal{F} \star \mathcal{G} = \{f \star g | f \in \mathcal{F}, g \in \mathcal{G}\}.$$

Lower (Definition 4.4 in [10]). *The lower Markov combination of $\mathcal{F}$ and $\mathcal{G}$ is defined as*

$$\mathcal{F} \underline{\star} \mathcal{G} = \{f \star g | f \in \mathcal{F}, g \in \mathcal{G}, f \text{ and } g \text{ consistent}\}.$$

Upper (Definition 4.5 in [10]). *The upper Markov combination of $\mathcal{F}$ and $\mathcal{G}$ is defined as*

$$\mathcal{F} \overline{\star} \mathcal{G} = \left\{ \frac{f\,g}{f_C}, \frac{f\,g}{g_C} \middle| f \in \mathcal{F}, g \in \mathcal{G} \right\}.$$

Super (Definition 4.7 and Proposition 4.14 in [10]). *The super Markov combination of $\mathcal{F}$ and $\mathcal{G}$ is defined as*

$$\mathcal{F} \circledast \mathcal{G} = \{f_{A|C} h_C g_{B|C} \text{ with } f \in \mathcal{F}, h \in \mathcal{F} \cup \mathcal{G} \text{ and } g \in \mathcal{G}\}.$$

These combinations are called Markov combinations because they are obtained by imposing the conditional independence relation $Y_A \perp\!\!\!\perp Y_B | Y_C$ (Proposition 4.3 of [10]). To simplify notation, we write $A \perp\!\!\!\perp B | C$ instead of $Y_A \perp\!\!\!\perp Y_B | Y_C$.

Clearly $\mathcal{F} \underline{\star} \mathcal{G} \subseteq \mathcal{F} \overline{\star} \mathcal{G} \subseteq \mathcal{F} \circledast \mathcal{G}$. Proposition 4.6 in [10] states that the lower Markov combination reduces or preserves the marginal families, namely $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(A\cup C)} \subseteq \mathcal{F}$ and $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(C\cup B)} \subseteq \mathcal{G}$. The reduction is strict when for example a distribution f in a family is not consistent with any element in the other family. The set

$$\mathcal{H} = \{h \text{ density over } A \cup C \cup B | h_{A\cup C} \in \mathcal{F}, h_{C\cup B} \in \mathcal{G} \text{ and } A \perp\!\!\!\perp B | C\}$$

equals $\mathcal{F} \underline{\star} \mathcal{G}$. Indeed, $\mathcal{F} \underline{\star} \mathcal{G} \subseteq \mathcal{H}$ follows from Propositions 4.1 and 4.3 in [10], while $\mathcal{H} \subseteq \mathcal{F} \underline{\star} \mathcal{G}$ follows from the fact that $h = h_{A\cup C} \star h_{C\cup B}$, for every $h \in \mathcal{H}$.

The upper Markov combination $\mathcal{F} \overline{\star} \mathcal{G}$ preserves or extends the original families whereas the super Markov combination is the largest set of distributions over $A \cup C \cup B$ which preserves or extends the original families. Furthermore, we have $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(A\cup C)} \subseteq \mathcal{F} \subseteq (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(A\cup C)}$ and $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(C\cup B)} \subseteq \mathcal{G} \subseteq (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(C\cup B)}$.

2.2. Gaussian independence models

A conditional independence statement imposes polynomial constraints on the elements of the variance-covariance matrix as shown in Proposition 2.1 below (see [12]).

Proposition 2.1. *Let A, B, C be disjoint subsets of $[n]$ and consider the non-degenerate Gaussian random vector $Y \sim \mathcal{N}_n(\mu, \Sigma)$. Then the following statements are equivalent*

1. *Y satisfies the conditional independence constraint $Y_A \perp\!\!\!\perp Y_B | Y_C$;*
2. *the sub-matrix $\Sigma_{A\cup C, B\cup C}$ has rank equal to the number of elements in C ($|C|$ in the following);*

3. *all size* $|C| + 1$ *minors of* $\Sigma_{A\cup C,B\cup C}$ *are zero*;
4. $\Sigma_{AB} = \Sigma_{AC}(\Sigma_{CC})^{-1}\Sigma_{CB}$.

When possible we write Σ_{AB} instead of $\Sigma_{A,B}$ for a matrix Σ and $A, B \subset [n]$.

The random vector Y satisfies the conditional independence statement $Y_A \perp\!\!\!\perp Y_B|Y_C$ if it satisfies simultaneously the inequality constraints expressed by $\Sigma \in \mathcal{S}_n^+$ and the polynomial constraints in item 3 of Proposition 2.1.

A popular class of conditional independence models is based on simple undirected graphs G with vertex set indexed by the elements of $[n]$. A graphical model with graph G gives a statistical model for the random vector $Y = (Y_i)_{i\in[n]}$ if and only if Y_i and Y_j are independent given $Y_{[n]\setminus\{i,j\}}$ whenever there is no edge between vertices i and j [6]. This is called the pairwise Markov property for graphical models. For non-degenerate Gaussian distributions (as those considered in this paper) it is equivalent to the local Markov property, the global Markov property and the factorisation of the distribution over the cliques of the graph. Hence, a Gaussian graphical model with graph G is defined by assigning a mean vector μ and a matrix $\Sigma \in \mathcal{S}_n^+$ where the entry (i, j) of Σ^{-1} is zero if and only if the edge (i, j) is not in the graph.

3. Matrix representation of Markov combinations for Gaussian distributions

In this section, we study the matrix representation of the Markov combination and its properties. In the following we assume a zero mean vector for all the Gaussian families under consideration. Let A, B, C be a partition of $[n]$. Consider two non-degenerate Gaussian random vectors, $Y_{A\cup C}$ defined on $A \cup C$ and $Y_{C\cup B}$ defined on $C \cup B$. Hence, $Y_{A\cup C} \sim \mathcal{N}_{|A\cup C|}(0, \Sigma)$ and $Y_{C\cup B} \sim \mathcal{N}_{|C\cup B|}(0, \Psi)$ with Σ and Ψ partitioned matrices as

$$\Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AC} \\ \Sigma_{CA} & \Sigma_{CC} \end{pmatrix}, \qquad \Psi = \begin{pmatrix} \Psi_{CC} & \Psi_{CB} \\ \Psi_{BC} & \Psi_{BB} \end{pmatrix},$$

where $(\Sigma_{CA})^T = \Sigma_{AC}$, $(\Psi_{BC})^T = \Psi_{CB}$ and ^{T} indicates transpose. Let $K = \Sigma^{-1}$ and $L = \Psi^{-1}$. In the following, f and g will denote the density of $Y_{A\cup C}$ and $Y_{C\cup B}$, respectively.

For $n_1, n_2 \subset [n]$ and a $|n_1| \times |n_2|$ matrix $M = (m_{ij})_{i\in n_1, j\in n_2}$, let $[M]^{[n]}$ be

$$[M]^{[n]} = \begin{cases} m_{ij}, & \text{if } i \in n_1, j \in n_2, \\ 0, & \text{otherwise.} \end{cases}$$

For example, $[\Psi]^{[n]}$ is a $n \times n$ matrix with zero entries in rows and columns indexed by elements in A and the elements of Ψ in $\{C \cup B\} \times \{C \cup B\}$. Its first $|A|$ rows and columns will be indexed by (elements in) A, subsequent rows and columns by C and then by B.

When f and g are Gaussian distributions, the operation $f \cdot g/g_C$ returns a Gaussian distribution with covariance and concentration matrices given in the following lemma which is a generalisation of Lemma 5.5 in [6].

Lemma 3.1. *The quantity $f \cdot g/g_C$ is a density of a Gaussian random vector $Y_{A \cup C \cup B} \sim \mathcal{N}_{|A \cup C \cup B|}(0, \Omega)$ with concentration matrix*, Ω^{-1},

$$\Omega^{-1} = \mathrm{Con}\left(\frac{f \cdot g}{g_C}\right) = [K]^{[n]} + [L]^{[n]} - \left[(\Psi_{CC})^{-1}\right]^{[n]} \tag{3.1}$$

and covariance matrix Ω

$$\Omega = \mathrm{Cov}\left(\frac{f \cdot g}{g_C}\right) = [\Sigma]^{[n]} + [\Psi]^{[n]} + [G]^{[n]} - [M]^{[n]}, \tag{3.2}$$

where

$$G = \begin{pmatrix} 0 & G_{AB} \\ G_{BA} & 0 \end{pmatrix}, \qquad M = \begin{pmatrix} M_{CC} & M_{CB} \\ M_{BC} & M_{BB} \end{pmatrix}$$

and

$$\begin{aligned} G_{AB} &= -\Sigma_{AC} L_{CB} (L_{BB})^{-1}, \\ M_{CB} &= (\Sigma_{CC} - \Psi_{CC}) L_{CB} (L_{BB})^{-1}, \\ M_{BB} &= -(L_{BB})^{-1} L_{BC} M_{CB}. \end{aligned}$$

Here G and M are block symmetric matrices, that is, $G_{BA} = (G_{AB})^T$, $M_{BC} = (M_{CB})^T$. *By construction $M_{CC} = \Psi_{CC}$.*

The matrix G_{AB} expresses the conditional independence constraint $A \perp\!\!\!\perp B | C$, M_{BB} can also be written as $M_{BB} = (L_{BB})^{-1} L_{BC} (\Psi_{CC} - \Sigma_{CC}) L_{CB} (L_{BB})^{-1}$ and for consistent distributions $M_{BB} = M_{CB} = 0$. A visually expressive re-writing of equations (3.1) and (3.2) is

$$\mathrm{Con}\left(\frac{f \cdot g}{g_C}\right) = \begin{pmatrix} K_{AA} & K_{AC} & 0 \\ K_{CA} & K_{CC} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & L_{CC} & L_{CB} \\ 0 & L_{BC} & L_{BB} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & (\Psi_{CC})^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{aligned} \mathrm{Cov}\left(\frac{f \cdot g}{g_C}\right) = {} & \begin{pmatrix} \Sigma_{AA} & \Sigma_{AC} & 0 \\ \Sigma_{CA} & \Sigma_{CC} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Psi_{CC} & \Psi_{CB} \\ 0 & \Psi_{BC} & \Psi_{BB} \end{pmatrix} + \begin{pmatrix} 0 & 0 & G_{AB} \\ 0 & 0 & 0 \\ (G_{AB})^T & 0 & 0 \end{pmatrix} \\ & - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Psi_{CC} & M_{CB} \\ 0 & (M_{CB})^T & M_{BB} \end{pmatrix}. \end{aligned}$$

In the literature, $f \cdot g/g_C$ is known as the operator of right combination and indicated as $f \triangleright g$, and $f \cdot g/f_C$ is known as the operator of left combination, defined as $f \triangleleft g := g \triangleright f$ (see [10] and references therein).

Lemma 3.1 provides the concentration and covariance matrices of the operator of right combination. They will be indicated in the rest of the paper as $\mathrm{Con}(f \triangleright g)$ and $\mathrm{Cov}(f \triangleright g)$, respectively. The concentration and covariance matrices of the operator of left combination, indicated as $\mathrm{Con}(g \rhd f)$ and $\mathrm{Cov}(g \rhd f)$, are obtained by the natural changes in Lemma 3.1. Equation (3.1), together with the facts that the conditional distribution of a non-degenerate Gaussian on some of its margins is still Gaussian and non-degenerate, and that sum of strictly positive definite matrices is strictly positive definite, shows that the right (left) combination of non-degenerate Gaussian distributions is non-degenerate.

If f and g are consistent, the operators of left and right combinations are equivalent to the Markov combination (see page 244 in [10]) and their covariance matrix follows from Lemma 3.1.

Proposition 3.1. *If f and g are consistent, that is, $\Sigma_{CC} = \Psi_{CC}$, the covariance matrix of the Markov combination is*

$$\mathrm{Cov}(f \star g) = |\Sigma|^{[n]} + |\Psi|^{[n]} + [G]^{[n]} - [\Psi_{CC}]^{[n]}$$

and the concentration matrix is

$$\mathrm{Con}(f \star g) = |K|^{[n]} + |L|^{[n]} - [\Psi_{CC}]^{[n]}.$$

Proposition 3.1 also gives the covariance and concentration matrices of the Markov combination when $C = \varnothing$. Under this constraint, f and g are consistent and independent and the concentration and covariance matrices of the Markov combination are respectively

$$\mathrm{Con}(f \star g) = \left[\Sigma^{-1}\right]^{[n]} + \left[\Psi^{-1}\right]^{[n]} \quad \text{and} \quad \mathrm{Cov}(f \star g) = |\Sigma|^{[n]} + |\Psi|^{[n]}.$$

A submatrix of the covariance matrix of the right combination is the covariance matrix of the distribution not appearing in the denominator. The same applies for the left combination. For the Markov combination, both initial matrices are submatrices of the covariance matrix of the combination. For left, right and Markov combinations, the concentration matrices have zeroes elements in the concentration matrices corresponding to the conditional independence $A \perp\!\!\!\perp B|C$. Indeed, the covariance matrix is filled with elements so that this constraint is satisfied, see also Proposition 4.3 in [10].

Example 3.1. Let $Y_{A\cup C} = \{Y_1, Y_2\}$ and $Y_{C\cup B} = \{Y_2, Y_3\}$ with $|A| = |B| = |C| = 1$, $\Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ and $\Psi = \begin{pmatrix} d & e \\ e & f \end{pmatrix}$.

(a) For $x = \frac{a}{ac-b^2} + \frac{f}{df-e^2} - \frac{1}{c}$, the covariance and concentration matrices for the left combination are

$$\mathrm{Cov}(g \rhd f) = \begin{pmatrix} a + \dfrac{b^2 d}{c^2} - \dfrac{b^2}{c} & db/c & eb/c \\ db/c & d & e \\ eb/c & e & f \end{pmatrix},$$

$$\mathrm{Con}(g \rhd f) = \begin{pmatrix} \dfrac{c}{ac-b^2} & \dfrac{-b}{ac-b^2} & 0 \\ \dfrac{-b}{ac-b^2} & x & \dfrac{-e}{df-e^2} \\ 0 & \dfrac{-e}{df-e^2} & \dfrac{d}{df-e^2} \end{pmatrix}.$$

The initial matrix Ψ is a sub-matrix of the covariance matrix of $\mathrm{Cov}(g \rhd f)$.

(b) Under the consistency assumption $c = d$, the combinations are all equal and their covariance matrix is

$$\mathrm{Cov}(f \star g) = \begin{pmatrix} a & b & eb/c \\ b & c & e \\ eb/c & e & f \end{pmatrix}.$$

The concentration matrix is as in (a) after substituting the consistency assumption. Both initial matrices Σ and Ψ are sub-matrices of $\mathrm{Cov}(f \star g)$.

Remark 3.1. The covariance matrix of the marginal distribution of any combination over a subset $D \subset [n]$ is

$$\mathrm{Cov}\big((f * g)^{\downarrow D}\big) = \big[\mathrm{Cov}(f * g)\big]_{DD},$$

where $*$ is any of $\{\lhd, \rhd, \star\}$. This follows from the usual results of the marginal distribution of a multivariate normal distribution and is particularly useful when dealing with super Markov combinations.

4. Matrix representation of Markov combinations for Gaussian families

In this section, we study the matrix representation of the lower, upper, super Markov combinations (see Definition 2.1) so that only matrix operations will be necessary to work quickly with the combinations.

Proposition 4.1. *Let A, B, C be a partition of $[n]$. Let $\mathcal{F}$ be a Gaussian model on $A \cup C$ and $\mathcal{G}$ on $C \cup B$. Let Σ (resp. Ψ) be a positive definite covariance matrix for $\mathcal{F}$ (resp. $\mathcal{G}$) and K (resp. L) its inverse.*

1. *The sets of concentration matrices $\mathrm{Con}(\mathcal{F} \underline{\star} \mathcal{G}) \in \mathcal{S}_n^+$ representing the lower Markov combinations and upper Markov combination of $\mathcal{F}$ and $\mathcal{G}$ are respectively,*

$$\mathrm{Con}(\mathcal{F} \underline{\star} \mathcal{G}) = \big\{[K]^{[n]} + [L]^{[n]} - \big[(\Psi_{CC})^{-1}\big]^{[n]} | \Psi_{CC} = \Sigma_{CC}, \Sigma \in \mathrm{Cov}(\mathcal{F}) \textit{ and } \Psi \in \mathrm{Cov}(\mathcal{G})\big\},$$

$$\begin{aligned}\mathrm{Con}(\mathcal{F} \overline{\star} \mathcal{G}) = {} & \big\{[K]^{[n]} + [L]^{[n]} - \big[(\Psi_{CC})^{-1}\big]^{[n]} | \Sigma \in \mathrm{Cov}(\mathcal{F}) \textit{ and } \Psi \in \mathrm{Cov}(\mathcal{G})\big\} \\ & \cup \big\{[K]^{[n]} + [L]^{[n]} - \big[(\Sigma_{CC})^{-1}\big]^{[n]} | \Sigma \in \mathrm{Cov}(\mathcal{F}) \textit{ and } \Psi \in \mathrm{Cov}(\mathcal{G})\big\}.\end{aligned}$$

2. *A matrix* $T \in \mathcal{S}_n^+$ *belongs to* $\operatorname{Cov}(\mathcal{F}\,\underline{\star}\,\mathcal{G})$ *if and only if*:
   (a) $T_{A\cup C,A\cup C} \in \operatorname{Cov}(\mathcal{F})$ *and* $T_{C\cup B,C\cup B} \in \operatorname{Cov}(\mathcal{G})$,
   (b) $T_{AB} = T_{AC}(T_{CC})^{-1}T_{CB}$.
3. *A matrix* $T \in \mathcal{S}_n^+$ *belongs to* $\operatorname{Cov}(\mathcal{F}\,\overline{\star}\,\mathcal{G})$ *if and only if*:
   (a) $T_{AB} = T_{AC}(T_{CC})^{-1}T_{CB}$,
   (b) $T_{A\cup C,A\cup C} \in \operatorname{Cov}(\mathcal{F})$ *and*
   (c) *there exists* $\Psi \in \operatorname{Cov}(\mathcal{G})$ *such that*

$$
\begin{aligned}
T_{CB} &= -(T_{CC} - \Psi_{CC})(\Psi_{CC})^{-1}\Psi_{CB},\\
T_{BB} &= \Psi_{BC}(\Psi_{CC})^{-1}(T_{CC} - \Psi_{CC})(\Psi_{CC})^{-1}\Psi_{CB}
\end{aligned}
$$

or

   (a) $T_{AB} = T_{AC}(T_{CC})^{-1}T_{CB}$,
   (b) $T_{C\cup B,C\cup B} \in \operatorname{Cov}(\mathcal{G})$ *and*
   (c) *there exists* $\Sigma \in \operatorname{Cov}(\mathcal{F})$ *such that*

$$
\begin{aligned}
T_{CA} &= -(T_{CC} - \Sigma_{CC})(\Sigma_{CC})^{-1}\Sigma_{CA},\\
T_{AA} &= \Sigma_{AC}(\Sigma_{CC})^{-1}(T_{CC} - \Sigma_{CC})(\Sigma_{CC})^{-1}\Sigma_{CA}.
\end{aligned}
$$

Proof. 1. Follows from the definition of lower and upper Markov combinations and by the generalisation of Lemma 3.1 to families of distributions.

2. Necessity: by definition of lower Markov combination and its properties. Sufficiency: items (2a) imply that there exists $f_1 \in \mathcal{F}$ and $g_1 \in \mathcal{G}$ with covariance matrices $T_{A\cup C,A\cup C}$ and $T_{C\cup B,C\cup B}$, respectively. This together with item (2a) guarantees that $f_1 \star g_1$ is the Markov combination and since $f_1 \star g_1 \subset \mathcal{F}\,\underline{\star}\,\mathcal{G}$ it follows that $\operatorname{Cov}(f_1 \star g_1) \in \operatorname{Cov}(\mathcal{F}\,\underline{\star}\,\mathcal{G})$. Also $\operatorname{Cov}(f_1 \star g_1) = T$ by Lemma 3.1.

3. The proof is similar to the one in 2 and is not reported here. □

Proposition 4.1 indicates how to compute the families of concentration and covariance matrices of lower and upper combinations by assembling concentration and covariance matrices of the original families. Only the submatrices Σ_{CC} and Ψ_{CC} related to the variables common to $\mathcal{F}$ and $\mathcal{G}$ need inverting.

Determining Markov combinations is equivalent to finding a positive definite completion of the covariance/concentration matrix defined over $A \cup C \cup B$. Such a completion is found imposing the constraint $A \perp\!\!\!\perp B|C$ that holds for all Markov combinations and the specific constraints imposed by each Markov combination. Lemma 3.1 and Proposition 4.1 show that the first constraint impacts the entries of the covariance/concentration matrices corresponding to variables in $A \cup B$ and the second constraint takes into account the covariance/concentration structure of the initial families over $A \cup C$ and $B \cup C$.

Remark 4.1. As every distribution in $\mathcal{F}\,\underline{\star}\,\mathcal{G}$ and $\mathcal{F}\,\overline{\star}\,\mathcal{G}$ is multivariate normal, the covariance matrix of the marginal distribution of any combination of Gaussian families of distributions over a subset $D \subset [n]$ is

$$\mathrm{Cov}\big((\mathcal{F} * \mathcal{G})^{\downarrow D}\big) = \big[\mathrm{Cov}(\mathcal{F} * \mathcal{G})\big]_{DD},$$

where $*$ is any of $\{\underline{\star}, \overline{\star}\}$.

Proposition 4.2 gives the concentration matrices of the super Markov combination. One way of defining it is via $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(A\cup C)} \overline{\star} (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(B\cup C)}$ (see Definition 4.7 in [10]) and therefore we can apply the formula for the upper Markov combination in Proposition 4.1 to the models $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(A\cup C)}$ and $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(C\cup B)}$.

Proposition 4.2. *The set of concentration matrices* $\mathrm{Con}(\mathcal{F} \otimes \mathcal{G}) \in \mathcal{S}_n^+$ *representing the super Markov combination of* $\mathcal{F}$ *and* $\mathcal{G}$ *is* $\mathrm{Con}(\mathcal{F} \otimes \mathcal{G}) = \mathrm{Con1} \cup \mathrm{Con2}$ *with*

$$\begin{aligned}
\mathrm{Con1} &= \big[\big(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})_{A\cup C, A\cup C}\big)^{-1}\big]^{[n]} + \big[\big(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})_{C\cup B, C\cup B}\big)^{-1}\big]^{[n]} \\
&\quad - \big[\big(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})_{A\cup C, A\cup C}\big)^{-1}_{CC}\big]^{[n]}, \\
\mathrm{Con2} &= \big[\big(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})_{A\cup C, A\cup C}\big)^{-1}\big]^{[n]} + \big[\big(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})_{C\cup B, C\cup B}\big)^{-1}\big]^{[n]} \\
&\quad - \big[\big(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})_{C\cup B, C\cup B}\big)^{-1}_{CC}\big]^{[n]}.
\end{aligned}$$

Proof. By Remark 4.1, $\mathrm{Cov}((\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(A\cup C)}) = [\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})]_{A\cup C, A\cup C}$ and $\mathrm{Cov}((\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow(C\cup B)}) = [\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G})]_{C\cup B, C\cup B}$. The thesis is obtained by using Proposition 4.1 and replacing K with $[(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G}))_{A\cup C, A\cup C}]^{-1}$, L with $[(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G}))_{C\cup B, C\cup B}]^{-1}$, Σ_{CC} with $[(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G}))_{A\cup C, A\cup C}]_{CC}$ and Ψ_{CC} with $[(\mathrm{Cov}(\mathcal{F} \overline{\star} \mathcal{G}))_{C\cup B, C\cup B}]_{CC}$. □

The following example highlights which elements of the initial covariance matrices are preserved in each Markov combination, which are completed by the conditional independence constraint $A \perp\!\!\!\perp B | C$ and which correspond to the constraints imposed by the combination.

Example 4.1 (Example 3.1 continued). Consider the Gaussian families $\mathcal{F}$ and $\mathcal{G}$ with covariances matrices Σ and Ψ as in Example 3.1, but this time Σ and Ψ are families of covariance matrices in $\mathcal{S}_2^+(\mathbb{R})$ that is the entries $a, b, c, \ldots$ take a number of values. Since for every $f \in \mathcal{F}$ there exist a $g \in \mathcal{G}$ such that $f_C = g_C$ (it is sufficient to take $c = d$) $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent. Hence, the lower Markov and meta Markov combinations are the same and are given by the family of distributions with covariance matrix given in part (c) of Example 3.1. Moreover, the covariance matrices of the lower-meta-upper-super-Markov combinations are all equal to

$$\mathrm{Cov}(\mathcal{F} \star \mathcal{G}) = \left\{ \begin{pmatrix} a & b & eb/c \\ b & c & e \\ eb/c & e & f \end{pmatrix} \right\}.$$

If $\mathcal{F}$ and $\mathcal{G}$ are such that $c \neq d$ for all $\Sigma \in \mathrm{Cov}(\mathcal{F})$ and $\Psi \in \mathrm{Cov}(\mathcal{G})$, then the families are not meta-consistent, the meta-Markov combination is not defined and the lower Markov combination is empty. The covariance matrices of the upper Markov combination are given by the set formed by the matrices for the left and right combinations of Example 3.1, i.e.,

$$\mathrm{Cov}(\mathcal{F}\,\bar{\star}\,\mathcal{G}) = \left\{ \begin{pmatrix} a + \dfrac{b^2 d}{c^2} - \dfrac{b^2}{c} & db/c & eb/c \\ db/c & d & e \\ eb/c & e & f \end{pmatrix}, \begin{pmatrix} a & b & be/d \\ b & c & ce/d \\ be/d & ce/d & f - \dfrac{e^2}{d} + \dfrac{e^2 c}{d^2} \end{pmatrix} \right\}.$$

The covariance matrices of $(\mathcal{F}\,\bar{\star}\,\mathcal{G})^{\downarrow(A\cup C)}$ and $(\mathcal{F}\,\bar{\star}\,\mathcal{G})^{\downarrow(C\cup B)}$ are

$$\mathrm{Cov}\big((\mathcal{F}\,\bar{\star}\,\mathcal{G})^{\downarrow(A\cup C)}\big) = \left\{ \begin{pmatrix} a + \dfrac{b^2 d}{c^2} - \dfrac{b^2}{c} & db/c \\ db/c & d \end{pmatrix}, \begin{pmatrix} a & b \\ b & c \end{pmatrix} \right\},$$

$$\mathrm{Cov}\big((\mathcal{F}\,\bar{\star}\,\mathcal{G})^{\downarrow(C\cup B)}\big) = \left\{ \begin{pmatrix} d & e \\ e & f \end{pmatrix}, \begin{pmatrix} c & ce/d \\ ce/d & f - \dfrac{e^2}{d} + \dfrac{e^2 c}{d^2} \end{pmatrix} \right\}.$$

This shows that $(\mathcal{F}\,\bar{\star}\,\mathcal{G})^{\downarrow(A\cup C)}$ and $(\mathcal{F}\,\bar{\star}\,\mathcal{G})^{\downarrow(C\cup B)}$ are meta-consistent as expected from Proposition 4.10 in [10] and that $\mathrm{Cov}(\mathcal{F} \otimes \mathcal{G}) = \mathrm{Cov}(\mathcal{F}\,\bar{\star}\,\mathcal{G})$.

Example 4.2. Two Gaussian families $\mathcal{F}$ and $\mathcal{G}$ where the lower Markov combination is the empty set and the super Markov combination is different from the upper Markov combination are $\mathcal{F} = \{f_1, f_2\}$ defined over $Y_{A\cup C} = \{Y_1, Y_2, Y_3\}$ and $\mathcal{G}$ defined over $Y_{C\cup B} = \{Y_2, Y_3, Y_4\}$ with

$$\mathrm{Cov}(f_1) = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 7 & \frac{3}{2} \\ 3 & \frac{3}{2} & 5 \end{pmatrix}, \qquad \mathrm{Cov}(f_2) = \begin{pmatrix} 6 & \frac{1}{2} & 5 \\ \frac{1}{2} & 1 & \frac{5}{12} \\ 5 & \frac{5}{12} & 10 \end{pmatrix} \quad \text{and} \quad \mathrm{Cov}(\mathcal{G}) = \begin{pmatrix} 8 & \frac{2}{3} & 2 \\ \frac{2}{3} & 2 & 3 \\ 2 & 3 & 9 \end{pmatrix}.$$

Variance and covariance matrices of the upper combination are found by suitable substitutions in Example 4.1.

5. Polynomial ideal representation of Markov combinations for Gaussian families

In this section, we study the Markov combinations with tools from algebraic statistics. We construct polynomial ideals for the Markov combinations from polynomial ideals of two original Gaussian families which admit polynomial representation. To make our aims clear, we remark that we wish to provide pointers to the usefulness of representing Gaussian models by polynomial ideals in order to study the properties of Markov combinations. For the full power of such a representation, we refer the interested reader to e.g. [4,7].

For Gaussian families conditional independence statements correspond to polynomial equality constraints on the entries of the covariance matrices. Inequality constraints are implicit in the positive definiteness of covariance matrices and correspond to imposing that all eigenvalues of the covariance matrices are positive. Equality constraints are mapped into polynomial ideals and the inequality constraints are imposed by intersecting with the set of positive definite matrices.

We adopt the following notation: $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$ indicates the set of all polynomials with real coefficients and n free variables (or indeterminates) $x_1, \ldots, x_n$ and $\mathbb{C}[x]$ when the coefficients are complex number. For $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$

$$\langle f_1, \ldots, f_k \rangle = \left\{ \sum_{i=1}^{k} s_i f_i : s_i \in \mathbb{K}[x], i = 1, \ldots, k \right\}$$

is the polynomial ideal generated by $f_1, \ldots, f_k \in \mathbb{K}[x]$; $\mathcal{V}(I) = \{v \in \mathbb{K}^n : f(v) = 0 \text{ for all } v \in I\}$ is the algebraic variety associated with a polynomial ideal $I \subset \mathbb{K}[x]$; and $I(\mathcal{V}) = \{f \in \mathbb{K}[x] : f(v) = 0 \text{ for all } v \in \mathcal{V}\}$ is the vanishing ideal associated to an algebraic variety $\mathcal{V}$. If $\mathcal{V}$ is not an algebraic variety, $I(\mathcal{V})$ can still be defined as above and its zero set is a superset of $\mathcal{V}$. For more background information on ideals and varieties, see [1] or other textbooks on algebraic geometry.

Let A, B, C be a partition of $[n]$. Proposition 2.1 states that the ideal corresponding to the variety $Y_A \perp\!\!\!\perp Y_B | Y_C$ is generated by the polynomials $\det(V)$ where V varies among all $(|C| + 1) \times (|C| + 1)$ submatrices of $\Sigma_{A\cup C, B\cup C}$ and det stands for determinant. Given a Gaussian family $\mathcal{F}$, we will denote with $I_{\mathcal{F}}$ the conditional independence ideal generated by $\mathcal{F}$ and with $\mathcal{V}(I_{\mathcal{F}}) \cap \mathcal{S}_n^+$ the variety of positive definite covariance matrices in $\mathcal{V}(I_{\mathcal{F}})$.

Not all Gaussian families $\mathcal{F}$ are algebraic varieties. Gaussian independence models, such as undirected graphical models, are a large class of statistical models which are algebraic varieties. This follows straightforwardly from Proposition 2.1.

5.1. Lower Markov combination

Let $\mathcal{F}$ (resp. $\mathcal{G}$) be a Gaussian (independence) model for $Y_{A\cup C} \sim \mathcal{N}_{|A\cup C|}(0, \Sigma)$ with $\Sigma \in \mathcal{S}^+_{|A\cup C|}$ (resp. $Y_{C\cup B} \sim \mathcal{N}_{|C\cup B|}(0, \Psi)$ with $\Psi \in \mathcal{S}^+_{|C\cup B|}$) and let $I_{\mathcal{F}}$ (resp. $I_{\mathcal{G}}$) be the corresponding (conditional independence) ideal. The set of covariance matrices of the Gaussian model $\mathcal{F}$ can be embedded in $\mathcal{S}_n^+$, simply by augmenting it with unknown parameters.

Example 5.1. For $\mathcal{F} = \{Y_2 \perp\!\!\!\perp Y_3 | Y_1\}$ and $n = 4$

$$\mathrm{Cov}(\mathcal{F})_3 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{12}\sigma_{13}/\sigma_{11} \\ \sigma_{13} & \sigma_{12}\sigma_{13}/\sigma_{11} & \sigma_{33} \end{pmatrix} \in \mathcal{S}_3^+$$

becomes

$$\mathrm{Cov}(\mathcal{F})_4 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{12}\sigma_{13}/\sigma_{11} & \sigma_{24} \\ \sigma_{13} & \sigma_{12}\sigma_{13}/\sigma_{11} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix} \in \mathcal{S}_4^+.$$

Ignoring inequality constraints, by Proposition 2.1 and the obvious lifting, $\mathrm{Cov}(\mathcal{F})_4$ can be seen as the variety corresponding to the ideal $\langle \sigma_{12}\sigma_{13} - \sigma_{23}\sigma_{11} \rangle$ in the polynomial ring with 10 variables $\mathbb{R}[\sigma_{ij} : 1 \le i \le j \le 4]$. Namely $\mathrm{Cov}(\mathcal{F})_4$ is the "cylinder" in a 10 dimensional space whose elements are $(a, \sigma_{14}, \sigma_{24}, \sigma_{34}, \sigma_{44})$ with $a = (a_i)_{i=1,\dots,6}$ such that

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ a_2 & a_4 & a_5 \\ a_3 & a_5 & a_6 \end{pmatrix} \in \mathrm{Cov}(\mathcal{F})_3$$

and $\sigma_{i4} \in \mathbb{R}$ for $i = 1, \dots, 4$. In the language of algebraic geometry, $\mathrm{Cov}(\mathcal{F})_4$ is the join between the variety $\sigma_{ij} = 0$ $(i, j = 1, 2, 3)$ and the variety

$$\begin{cases} \sigma_{i4} = 0, & i = 1, 2, 3, 4, \\ \sigma_{12}\sigma_{13} - \sigma_{23}\sigma_{11} = 0, \end{cases}$$

where the join of two algebraic varieties is the smallest algebraic variety containing both.

With slight abuse of notation we call $I_{\mathcal{F}}$ the ideal both in the smaller and larger set of variables, $\mathbb{R}[\sigma_{ij} : 1 \le i \le j \le |A|]$ and $\mathbb{R}[\sigma_{ij} : 1 \le i \le j \le n]$ respectively, as the two ideals are generated by the same polynomials.

We showed in Section 2.1 that $\mathcal{F} \underline{\star} \mathcal{G}$ is equal to

$$\mathcal{H} = \{h \text{ density over } A \cup C \cup B \mid h_{A\cup C} \in \mathcal{F}, h_{C\cup B} \in \mathcal{G} \text{ and } A \perp\!\!\!\perp B \mid C\}$$

and Proposition 4.1 substantially identifies the structure of $\mathcal{H}$ suggesting to intersect $\mathrm{Cov}(\mathcal{F})_n$ with $\mathrm{Cov}(\mathcal{G})_n$ and with the variety corresponding to $A \perp\!\!\!\perp B|C$. But intersection of algebraic varieties corresponds to (the variety of the) sum of ideals. Hence, we have proven that:

Proposition 5.1. *The polynomial ideal representation of* $\mathcal{F} \underline{\star} \mathcal{G}$ *is*

$$I_{\mathcal{F}\underline{\star}\mathcal{G}} = I_{\mathcal{F}} + I_{\mathcal{G}} + I_{A\perp\!\!\!\perp B|C} \subset \mathbb{R}[\sigma_{ij} : 1 \le i \le j \le n]$$

and the closure (*in the Zariski topology*) *of* $\mathcal{F} \underline{\star} \mathcal{G}$ *is*

$$\mathcal{V}(I_{\mathcal{F}\underline{\star}\mathcal{G}}) = \mathcal{V}(I_{\mathcal{F}}) \cap \mathcal{V}(I_{\mathcal{G}}) \cap \mathcal{V}(I_{A\perp\!\!\!\perp B|C}).$$

The construction above embodies the fact that $\mathrm{Cov}(\mathcal{F})_n$ and $\mathrm{Cov}(\mathcal{G})_n$ are two "cylinders" in $\mathbb{R}^{n(n+1)/2}$ with coordinates σ_{ij}, $1 = i \le j \le n$, orthogonal to the non-common variables with respect to the standard scalar product in $\mathbb{R}^n$. Their intersection corresponds to consistent distributions and in turn it has to be intersected with the variety corresponding to the conditional independence statement $A \perp\!\!\!\perp B|C$.

The following examples can be checked with a computer algebra software, e.g. the *PolynomialIdeals* package in Maple [8]. Computations among ideals are performed over the complex field and then intersection with the real numbers is applied.

The Gaussian families $\mathcal{F}$ and $\mathcal{G}$ are written by means of the conditional independence relations between the random variables. The conditional independence relations are written in terms of the

Figure 1. From left to right, the two Gaussian graphical models $\mathcal{F} = \{Y_2 \perp\!\!\!\perp Y_3|Y_1\}$ and $\mathcal{G} = \{Y_2 \perp\!\!\!\perp Y_3\}$.

matrices of the covariance matrix of the Gaussian distributions associated with the combination. For simplicity, the elements of $\mathrm{Cov}(\mathcal{F} \underline{\star} \mathcal{G})$ will be indicated as $\{\sigma_{ij}\}$ with $1 \leq i \leq j \leq n$. Of course, as we have seen in the previous sections, they belong to either $\mathrm{Cov}(\mathcal{F})$ or $\mathrm{Cov}(\mathcal{G})$ and for a non-empty lower Markov combination we need consistency over the elements belonging to both $\mathrm{Cov}(\mathcal{F})$ and $\mathrm{Cov}(\mathcal{G})$.

The ideal generated by zero, which is the identity for the operation sum-of-ideals, is the conditional independence ideal of a saturated model, meaning that no polynomial constraint is imposed on the entries of the model covariance matrices.

Example 5.2. The conditional independence ideals of the Gaussian families $\mathcal{F} = \{Y_2 \perp\!\!\!\perp Y_3|Y_1\}$ and $\mathcal{G} = \{Y_2 \perp\!\!\!\perp Y_3\}$ in Figure 1 are

$$I_{\mathcal{F}} = I_{Y_2 \perp\!\!\!\perp Y_3|Y_1} = \langle \sigma_{23}\sigma_{11} - \sigma_{12}\sigma_{13} \rangle \quad \text{and} \quad I_{\mathcal{G}} = I_{Y_2 \perp\!\!\!\perp Y_3} = \langle \sigma_{23} \rangle.$$

The polynomial ideal representation of $\mathcal{F} \underline{\star} \mathcal{G}$ is

$$I_{\mathcal{F} \underline{\star} \mathcal{G}} = I_{\mathcal{F}} + I_{\mathcal{G}} + \langle 0 \rangle = I_{Y_2 \perp\!\!\!\perp Y_3|Y_1} + I_{Y_2 \perp\!\!\!\perp Y_3} = \langle \sigma_{23}\sigma_{11} - \sigma_{12}\sigma_{13}, \sigma_{23} \rangle.$$

The primary decomposition of $I_{\mathcal{F} \underline{\star} \mathcal{G}}$ is

$$I_{\mathcal{F} \underline{\star} \mathcal{G}} = \langle \sigma_{23}, \sigma_{12} \rangle \cap \langle \sigma_{23}, \sigma_{13} \rangle = I_{Y_2 \perp\!\!\!\perp Y_{\{1,3\}}} \cap I_{Y_3 \perp\!\!\!\perp Y_{\{1,2\}}}.$$

The Zariski closure of $\mathcal{F} \underline{\star} \mathcal{G}$ is then

$$\mathcal{V}(I_{\mathcal{F} \underline{\star} \mathcal{G}}) = \mathcal{V}(I_{Y_2 \perp\!\!\!\perp Y_3|Y_1} \cap I_{Y_2 \perp\!\!\!\perp Y_3}) = \mathcal{V}(I_{Y_2 \perp\!\!\!\perp Y_{\{1,3\}}}) \cup \mathcal{V}(I_{Y_3 \perp\!\!\!\perp Y_{\{1,2\}}}).$$

This shows that $\mathcal{F} \underline{\star} \mathcal{G}$ is the union of two Gaussian graphical models given by $Y_{\{1,3\}} \perp\!\!\!\perp Y_2$ and $Y_{\{1,2\}} \perp\!\!\!\perp Y_3$. In particular, $\mathcal{F} \underline{\star} \mathcal{G}$ is not a graphical combination, that is, $\mathcal{F} \underline{\star} \mathcal{G}$ is not a graphical model.

The singular locus of $\mathcal{V}(I_{\mathcal{F} \underline{\star} \mathcal{G}})$, intuitively the points where the tangent line fails to exist, can now be computed as

$$\mathcal{V}(I_{Y_2 \perp\!\!\!\perp Y_{\{1,3\}}}) \cap \mathcal{V}(I_{Y_3 \perp\!\!\!\perp Y_{\{1,2\}}}) = \{\sigma_{12} = \sigma_{13} = \sigma_{23} = 0\}$$

giving the complete independence model $Y_1 \perp\!\!\!\perp Y_2 \perp\!\!\!\perp Y_3$. For the usefulness of this type computation in, for example, likelihood ratio tests see [4].

Example 5.3. The three saturated Gaussian families for (Y_1, Y_2) (say $\mathcal{F}$), for (Y_2, Y_3) (say $\mathcal{G}$) and for (Y_3, Y_4) (say $\mathcal{K}$) do not satisfy the conditions for associativity given in [10], page 244. Non associativity can be shown algebraically in a straightforward manner. According to Proposition 5.1 and recalling that $I_{\mathcal{F}} = I_{\mathcal{G}} = I_{\mathcal{K}} = \langle 0 \rangle$, we need to compare the varieties of the two ideals

$$\begin{aligned}
I_{(\mathcal{F}\underline{\star}\mathcal{G})\underline{\star}\mathcal{K}} &= (I_{Y_1 \perp\!\!\!\perp Y_3|Y_2}) + I_{Y_4 \perp\!\!\!\perp (Y_1,Y_2)|Y_3} \\
&= \langle \underline{\sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{22}}, \sigma_{13}\sigma_{24} - \sigma_{23}\sigma_{14}, \sigma_{13}\sigma_{34} - \sigma_{33}\sigma_{14}, \underline{\sigma_{23}\sigma_{34} - \sigma_{33}\sigma_{24}} \rangle, \\
I_{\mathcal{F}\underline{\star}(\mathcal{G}\underline{\star}\mathcal{K})} &= I_{Y_1 \perp\!\!\!\perp (Y_3,Y_4)|Y_2} + (I_{Y_2 \perp\!\!\!\perp Y_4|Y_3}) \\
&= \langle \underline{\sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{22}}, \sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22}, \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{33}, \underline{\sigma_{23}\sigma_{34} - \sigma_{33}\sigma_{24}} \rangle.
\end{aligned}$$

They are radical ideals and are different, hence so are their varieties. Indeed the underlined polynomials are in both ideals, but the polynomial constraint $\sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22}$ must be satisfied by the entries of covariance matrices for $\mathcal{F}\underline{\star}(\mathcal{G}\underline{\star}\mathcal{K})$ and is not satisfied by models in $(\mathcal{F}\underline{\star}\mathcal{G})\underline{\star}\mathcal{K}$. Indeed the normal form [1] of $\sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22}$ with respect to $I_{(\mathcal{F}\underline{\star}\mathcal{G})\underline{\star}\mathcal{K}}$ is not zero. In conclusion $(\mathcal{F}\underline{\star}\mathcal{G})\underline{\star}\mathcal{K} \neq \mathcal{F}\underline{\star}(\mathcal{G}\underline{\star}\mathcal{K})$.

Example 5.4. The ideal of the lower Markov combination of the non-graphical model $\mathcal{H}$ given by $Y_{\{1,3\}} \perp\!\!\!\perp Y_2$ and $Y_{\{1,2\}} \perp\!\!\!\perp Y_3$, obtained in Example 5.2, with the graphical model $\mathcal{K} = \{Y_2 \perp\!\!\!\perp Y_3|Y_4\}$ is given by

$$\begin{aligned}
I_{\mathcal{H}\underline{\star}\mathcal{K}} &= I_{\mathcal{F}\underline{\star}\mathcal{G}} + I_{\mathcal{K}} + I_{Y_1 \perp\!\!\!\perp Y_4|Y_{\{2,3\}}} \\
&= I_{\mathcal{F}\underline{\star}(\mathcal{G}\underline{\star}\mathcal{K})}.
\end{aligned}$$

This is the particular case discussed in [10], page 244, where associativity of Markov combinations holds. We have

$$\begin{aligned}
I_{\mathcal{H}\underline{\star}\mathcal{K}} &= \langle \sigma_{23}\sigma_{11} - \sigma_{12}\sigma_{13}, \sigma_{23} \rangle + \langle \sigma_{23}\sigma_{44} - \sigma_{24}\sigma_{34} \rangle \\
&\quad + \big\langle \sigma_{23}(\sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23}) - \sigma_{22}(\sigma_{13}\sigma_{34} - \sigma_{14}\sigma_{33}) + \sigma_{12}(\sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{33}) \big\rangle \\
&= \big\langle \sigma_{23}, \sigma_{12}\sigma_{13}, \sigma_{24}\sigma_{34}, \sigma_{22}(\sigma_{13}\sigma_{34} - \sigma_{14}\sigma_{33}) + \sigma_{12}\sigma_{24}\sigma_{33} \big\rangle.
\end{aligned}$$

As $\sigma_{ii} \neq 0, i = 1, \dots, 4$ by equating to zero simultaneously the above generators of $I_{\mathcal{H}\underline{\star}\mathcal{K}}$ we find that in the lower Markov combination $\mathcal{H}\underline{\star}\mathcal{K}$ there are four types of symmetric matrices each one of which is defined by one of the following lines

$$\begin{aligned}
&0 = \sigma_{23} = \sigma_{12} = \sigma_{24} = \sigma_{13}\sigma_{34} - \sigma_{14}\sigma_{33}, \\
&0 = \sigma_{23} = \sigma_{12} = \sigma_{14} = \sigma_{34}, \\
&0 = \sigma_{23} = \sigma_{13} = \sigma_{14} = \sigma_{24}, \\
&0 = \sigma_{23} = \sigma_{13} = \sigma_{34} = \sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22}.
\end{aligned}$$

5.2. Upper and super Markov combinations

Proposition 5.2. *The polynomial ideal representation of* $\mathcal{F}\overline{\star}\mathcal{G}$ *is*

$$I_{\mathcal{F}\overline{\star}\mathcal{G}}=(I_{\mathcal{F}}\cap I_{\mathcal{G}})+I_{A\perp\!\!\!\perp B|C}\subset\mathbb{R}[\sigma_{ij}:1\le i\le j\le n]$$

and its closure (*in the Zariski topology*) *is*

$$\mathcal{V}(I_{\mathcal{F}\overline{\star}\mathcal{G}})=\bigl(\mathcal{V}(I_{\mathcal{F}})\cup\mathcal{V}(I_{\mathcal{G}})\bigr)\cap\mathcal{V}(I_{A\perp\!\!\!\perp B|C}).$$

Proof. The result is proven with similar arguments of Proposition 5.1. The densities in $\mathcal{F}\overline{\star}\mathcal{G}$ satisfy the conditional independence statement and their marginal belong to $\mathcal{F}$ or $\mathcal{G}$, thus identifying $\mathcal{F}\overline{\star}\mathcal{G}$ with

$$\mathcal{K}=\{h\text{ density over }A\cup B|h_A\in\mathcal{F}\text{ or }h_B\in\mathcal{G}\text{ and }A\perp\!\!\!\perp B|C\}.$$ □

Proposition 5.3. *The polynomial ideal representation of* $\mathcal{F}\otimes\mathcal{G}$ *is*

$$I_{\mathcal{F}\otimes\mathcal{G}}=(I_{A\cup C}\cap I_{C\cup B})+I_{A\perp\!\!\!\perp B|C}\subset\mathbb{R}[\sigma_{ij}:1\le i\le j\le n]$$

where $I_{A\cup C}=\sqrt{I_{\mathcal{F}\overline{\star}\mathcal{G}}\cap\mathbb{C}[x_A]}$ *and* $I_{C\cup B}=\sqrt{I_{\mathcal{F}\overline{\star}\mathcal{G}}\cap\mathbb{C}[x_B]}$ *are elimination ideals of the variables in* B *and in* A, *respectively. Also*, *the closure* (*in the Zariski topology*) *of* $\mathcal{F}\otimes\mathcal{G}$ *is*

$$\mathcal{V}(I_{\mathcal{F}\overline{\star}\mathcal{G}})=\bigl(\mathcal{V}(I_{A\cup C})\cup\mathcal{V}(I_{C\cup B})\bigr)\cap\mathcal{V}(I_{A\perp\!\!\!\perp B|C}).$$

Proof. We already observed that $\mathcal{F}\otimes\mathcal{G}=(\mathcal{F}\overline{\star}\mathcal{G})^{\downarrow(A\cup C)}\overline{\star}(\mathcal{F}\overline{\star}\mathcal{G})^{\downarrow(B\cup C)}$. The polynomial representation of $(\mathcal{F}\overline{\star}\mathcal{G})^{\downarrow(A\cup C)}$ is given by $I_{A\cup C}$ (see, e.g., [12]). The ideal representation of the super Markov combination then follows from Proposition 5.2. □

From a computational view point Proposition 5.3 is less useful than the two analogue propositions for lower and upper Markov combinations because it involves elimination ideals which are better computed over the algebraically closed field $\mathbb{C}$ (see [1], Chapter 3).

Example 5.5 (Example 5.2 continued). The polynomial ideal for $\mathcal{F}\overline{\star}\mathcal{G}$ is

$$I_{\mathcal{F}\overline{\star}\mathcal{G}}=I_{\mathcal{F}}\cap I_{\mathcal{G}}=I_{Y_2\perp\!\!\!\perp Y_3|Y_1}\cap I_{Y_2\perp\!\!\!\perp Y_3}=\bigl\langle\sigma_{23}(\sigma_{23}\sigma_{11}-\sigma_{12}\sigma_{13})\bigr\rangle.$$

Its Zariski closure is a non-graphical Gaussian model. Comparison with Example 5.2 gives a clear example of the inclusion reversing relationship between ideals and varieties: $\mathcal{F}\underline{\star}\mathcal{G}\subset\mathcal{F}\overline{\star}\mathcal{G}$ while $I_{\mathcal{F}\underline{\star}\mathcal{G}}\supset I_{\mathcal{F}\overline{\star}\mathcal{G}}$. In particular, $\mathcal{F}\overline{\star}\mathcal{G}$ contains Gaussian densities for which $Y_2\perp\!\!\!\perp Y_3$ ($\sigma_{23}=0$) *or* $Y_2\perp\!\!\!\perp Y_3|Y_1$ ($\sigma_{23}\sigma_{11}-\sigma_{12}\sigma_{13}=0$), while $\mathcal{F}\underline{\star}\mathcal{G}$ contains Gaussian densities for which $Y_2\perp\!\!\!\perp Y_3$ *and* $Y_2\perp\!\!\!\perp Y_3|Y_1$. The polynomial ideal representation of $\mathcal{F}\otimes\mathcal{G}$ is

$$I_{\mathcal{F}\otimes\mathcal{G}}=(I_{A\cup C}\cap I_{C\cup B})+I_{A\perp\!\!\!\perp B|C}=\langle 0\rangle,$$

and its Zariski closure is a saturated Gaussian graphical model.

6. An illustrative example

In [2], the covariance between variables that are not jointly observed is estimated via a factor analysis model. A sample of women ($n = 25\,118$) were examined in a test composed by 5 sections: two maths and two verbal sections of the Scholastic Aptitude Test followed by one section of the Test of Standard Written English (TSWE in the following) denoted as M1, M2, V1, V2, T1, respectively. A subsample ($n_1 = 12\,761$) were given Section 2 of the TSWE, and another subsample ($n_2 = 12\,357$) were given Section 3. They are denoted as T2 and T3, respectively. No examinee was given both T2 and T3. The covariance matrix of (M1, M2, V1, V2, T1, T2), called T2F, and the covariance matrix of (M1, M2, V1, V2, T1, T3), called T3F, are shown in Table 1 together with their correlations (in the upper diagonal part of the matrices). They are taken from Table 6 (sections (b) and (d), respectively) of [5]. The estimate of $\mathrm{Cov}(\mathrm{T2}, \mathrm{T3})$ is needed to equate tests based on these sections.

We estimate the joint covariance matrix of (M1, M2, V1, V2, T1, T2, T3) using the Markov combinations. Table 2 shows the upper Markov combination (identical to the super Markov combination) of T2F and T3F. The values of $\mathrm{Cov}(\mathrm{T2}, \mathrm{T3})$ of the left and right combinations are very similar (66.99 and 66.47, respectively) and not too dissimilar from the estimate of [2] which is given as 72.71 (see Table 3 of [2]). In [2], the result is obtained by assuming a factor analysis joint model of (M1, M2, V1, V2, T1, T2, T3) and three correlated hidden factors.

Suppose now we are interested in investigating a different scenario based on this framework. Let us impose some constraints on T2F and T3F to reflect the fact that certain blocks of T2F and T3F contain very similar covariances (and correlations) which can be assumed equal (see Table 1). The polynomial representation of the constrained Gaussian models are given by the

Table 1. (a) Covariance and correlation matrices for T2F data; (b) covariance and correlation matrices for T3F data

	M1	M2	V1	V2	T1	T2
(a) T2F						
M1	32.47	0.813	0.638	0.639	0.627	0.610
M2	39.23	51.85	0.625	0.628	0.629	0.619
V1	31.66	39.23	75.85	0.838	0.764	0.737
V2	26.67	33.12	53.41	53.53	0.768	0.738
T1	37.25	47.26	69.41	58.62	108.63	0.861
T2	33.35	42.78	61.50	51.83	86.07	91.90
(b) T3F						
M1	33.59	0.814	0.640	0.649	0.632	0.601
M2	34.58	53.70	0.628	0.638	0.637	0.607
V1	32.49	40.33	76.74	0.841	0.772	0.741
V2	27.76	34.54	54.39	54.48	0.772	0.749
T1	38.17	48.62	70.46	59.37	108.53	0.859
T3	31.81	40.63	59.36	50.55	81.79	83.49

Table 2. Covariance matrices obtained via the upper Markov combination (identical results for the super Markov combination). (a) Covariance matrix corresponding to the operator of left combination; (b) covariance matrix corresponding to the operator of right combination

	M1	M2	V1	V2	T1	T2	T3
(a) Left combination							
M1	33.59						
M2	34.58	53.70					
V1	32.49	40.33	76.74				
V2	27.76	34.54	54.39	54.48			
T1	38.17	48.62	70.46	59.37	108.53		
T2	33.07	44.89	62.48	52.66	86.38	92.77	
T3	31.81	40.63	59.36	50.55	81.79	66.99	83.49
(b) Right combination							
M1	32.47						
M2	39.23	51.85					
V1	31.66	39.23	75.85				
V2	26.67	33.12	53.41	53.53			
T1	37.25	47.26	69.41	58.62	108.63		
T2	33.35	42.78	61.50	51.83	86.07	91.90	
T3	31.23	39.48	58.44	49.78	81.54	66.47	83.09

ideals

$$I_{\mathrm{T2F}} = \langle \sigma_{1k} = \sigma_{2k} = c, \sigma_{3l} = \sigma_{4l} = d \rangle, \qquad I_{\mathrm{T3F}} = \langle \sigma_{1m} = \sigma_{2m} = c, \sigma_{3n} = \sigma_{4n} = d \rangle,$$

with $k = \{3, 4, 5, 6\}$ and $l = \{5, 6\}$, $m = \{3, 4, 5, 7\}$, $n = \{5, 7\}$. The parameter σ_{11} corresponds to the variance on M_1, σ_{66} to the variance on T_2 and σ_{77} to the variance on T_3. The meaning of all other σ's follows now easily. Inequalities constraints might be added by further assuming $\sigma_{ij} > 0$, $i, j = \{1, \ldots, 7\}$. The two ideals of the lower and upper Markov combinations are

$$(I_{\mathrm{T2F}} + I_{\mathrm{T3F}}) + I_{T_2 \perp\!\!\!\perp T_3 | (M_1, M_2, V_1, V_2, T_1)} \quad \text{and} \quad (I_{\mathrm{T2F}} \cap I_{\mathrm{T3F}}) + I_{T_2 \perp\!\!\!\perp T_3 | (M_1, M_2, V_1, V_2, T_1)},$$

respectively. Here $I_{T_2 \perp\!\!\!\perp T_3 | (M_1, M_2, V_1, V_2, T_1)}$ is the ideal generated by the determinant of $\Sigma_{A\cup C, B\cup C}$ with $A = \{T_2\}$, $B = \{T_3\}$, $C = \{M_1, M_2, V_1, V_2, T_1\}$.

Inspection of a generating set of these ideals shows that the stationarity constraints in both Gaussian families act on the homogeneous polynomial generated by imposing $T_2 \perp\!\!\!\perp T_3 | (M_1, M_2, V_1, V_2, T_1)$. Furthermore any 7×7 matrix of the lower Markov combination has $\sigma_{1k} = \sigma_{2k} = c$ for $k = \{3, 4, 5, 6, 7\}$ and $\sigma_{3l} = \sigma_{4l} = d$ for $l = \{5, 6, 7\}$. Any matrix of the upper combination has $\sigma_{1k} = \sigma_{2k} = c$ for $k = \{3, 4, 5\}$ and $\sigma_{3l} = \sigma_{4l} = d$ for $l = 5$. In the upper Markov combinations there are models obtained from the lower Markov combination by relaxing any subset of the four constraints for $k, l = 6$ (and by symmetry also $k, l = 7$). Finally from the structure of the polynomial ideals of both combinations, we can see that they not depend on the variances of T2 and T3.

All the necessary computations have been carried out in Maple [8] as in previous examples. With more refined computations the singularities and other properties of the models can be elucidated.

Knowing the algebraic structure and properties of the statistical models of the combinations could help in determining the right assumptions to be made and in choosing the appropriate combination in different applied contexts.

7. Conclusion

This paper provides two algebraic representations of the Markov combinations for structural meta-analysis problems described in [10]. Lemma 3.1 and Proposition 4.1 give the covariance and concentration matrices for the different types of combinations in terms of the entries of the covariance and concentration matrices of the models to be combined, for Gaussian distributions and for Gaussian statistical models, respectively. The Markov combinations set to zero the entries of the concentration matrices corresponding to variables not jointly observed. This expresses conditional independence between these variables and is a particular case of matrix completion.

An algebraic/geometric setting can be useful in studying the constraints imposed by each initial model and those arising in the joint model. Proposition 5.1 describes how to construct the polynomial ideal for the lower Markov combination starting from the ideals of the models to be combined. similar reasoning to that adopted in Proposition 5.1 leads to the polynomial representations of upper and super Markov combinations in Section 5.2. Examples show uses of the obtained independence ideals to derive information on the corresponding statistical models. These uses have the same computational limits as the available softwares for handling polynomials.

Besides the independence constraints related to Markov combinations, other types of polynomial constraints can be imposed on a Gaussian statistical model as shown in Section 6 where equality constraints have been imposed. This paper gives only pointers to the usefulness of representing Gaussian models by polynomial ideals and it is besides its aim to show the full power of such representation for which we refer to for example, [4,7].

Acknowledgements

The authors are grateful to the anonymous referees and the Associate Editor whose feedback helped to improve the paper. The authors would also like to thank Professor Aldo Conca, Professor Steffen Lauritzen and Sir David Cox for their comments on early drafts of the paper.

References

[1] Cox, D., Little, J. and O'Shea, D. (1997). *Ideals, Varieties, and Algorithms*, 2nd ed. *Undergraduate Texts in Mathematics*. New York: Springer. An introduction to computational algebraic geometry and commutative algebra. MR1417938

[2] Cudeck, R. (2000). An estimate of the covariance between variables which are not jointly observed. *Psychometrika* **65** 539–546. MR1849281

[3] Dawid, A.P. and Lauritzen, S.L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann*. *Statist*. **21** 1272–1317. MR1241267

[4] Drton, M. (2009). Likelihood ratio tests and singularities. *Ann*. *Statist*. **37** 979–1012. MR2502658

[5] Holland, P.W. and Wrightman, L.E. (1982). Section pre-equating: A preliminary investigation. In *Test Equating* (P.W. Holland and D.R. Rubin, eds.) 271–297. New York: Academic Press.

[6] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: The Clarendon Press, Oxford Univ. Press. MR1419991

[7] Lin, S., Uhler, C., Sturmfels, B. and Bühlmann, P. (2014). Hypersurfaces and their singularities in partial correlation testing. *Found*. *Comput*. *Math*. **14** 1079–1116. MR3260260

[8] Maple 15 (2011). Maplesoft, a division of Waterloo Maple Inc. Waterloo, Ontario.

[9] Massa, M.S. and Chiogna, M. (2013). Effectiveness of combinations of Gaussian graphical models for model building. *J*. *Stat*. *Comput*. *Simul*. **83** 1602–1612. MR3169259

[10] Massa, M.S. and Lauritzen, S.L. (2010). Combining statistical models. In *Algebraic Methods in Statistics and Probability II*. *Contemp*. *Math*. **516** 239–259. Providence, RI: Amer. Math. Soc. MR2730753

[11] Pistone, G., Riccomagno, E. and Wynn, H.P. (2001). *Algebraic Statistics*: *Computational Commutative Algebra in Statistics*. *Monographs on Statistics and Applied Probability* **89**. Boca Raton, FL: Chapman & Hall/CRC. MR2332740

[12] Sullivant, S. (2008). Algebraic geometry of Gaussian Bayesian networks. *Adv. in Appl*. *Math*. **40** 482–513. MR2412156

Received November 2013 and revised August 2015