

Wishart distributions for decomposable graphs

G erard Letac*, H el ene Massam†

Abstract

When considering a graphical Gaussian model \mathcal{N}_G Markov with respect to a decomposable graph G , the parameter space of interest for the precision parameter is the cone P_G of positive definite matrices with fixed zeros corresponding to the missing edges of G . The parameter space for the scale parameter of \mathcal{N}_G is the cone Q_G , dual to P_G , of incomplete matrices with submatrices corresponding to the cliques of G being positive definite. In this paper we construct on the cones Q_G and P_G two families of Wishart distributions, namely the type I and type II Wisharts. They can be viewed as a generalization of the hyper Wishart and the inverse of the hyper inverse Wishart as defined by Dawid and Lauritzen (1993). We will show that the type I and II Wisharts have properties similar to those of the hyper and hyper inverse Wishart. Indeed, the inverse of the type II Wishart forms a conjugate family of priors for the covariance parameter of the graphical Gaussian model and is strong directed hyper Markov for every direction given to the graph by a perfect order of its cliques, while the type I Wishart is weak hyper Markov. Moreover, the inverse type II Wishart as a conjugate family presents the advantage of having a multi-dimensional shape parameter, thus offering flexibility for the choice of a prior.

Both types I and II Wishart distributions depend on multivariate shape parameters. A shape parameter is acceptable if and only if it satisfies a certain eigenvalue property. We show that the sets of acceptable shape parameters for a non complete G have dimension at least equal to one plus the number of cliques in G . These families, as conjugate families, are richer than the traditional Diaconis-Ylvisaker conjugate families which all have a shape parameter set of dimension one. A decomposable graph which does not contain a three-link chain as an induced subgraph is said to be homogeneous. In this case, our Wisharts are particular cases of the Wisharts on homogeneous cones as defined by Andersson and Wojnar (2004) and the dimension of the shape parameter set is even larger than in the non homogeneous case: it is indeed equal to the number of cliques plus the number of distinct minimal separators. Using the model where G is a three-link chain, we show by computing a 7-uple integral that in general we cannot

*Laboratoire de Statistique et Probabilit es, Universit e Paul Sabatier, 31062, Toulouse, France

†Department of Statistics, York University, Toronto, M3J 1P3, Canada. This author was supported by NSERC grant A 8947.

expect the shape parameter sets to have dimension larger than the number of cliques plus one.

Keywords: graphical models, covariance selection models, hyper Markov, conjugate priors, perfect order, triangulated graphs, chordal graphs, inverse Wishart, hyper Wishart and hyper inverse Wishart, homogeneous cones, natural exponential families.

AMS 2000 Subject classifications. Primary 62H99; Secondary 62E15

1 Introduction

The primary aim of this paper is to develop a new family of conjugate prior distributions with attractive Markov properties for the covariance parameter, or equivalently the precision parameter, of graphical Gaussian models Markov with respect to a decomposable graph G . While doing so, we are led to define two new classes of Wishart distributions and their inverses and to study their properties.

Let us recall that an undirected graph is a pair (V, \mathcal{E}) where $V = \{1, \dots, r\}$ and \mathcal{E} is a family of subsets $\{i, j\}$ of V , of size 2. It will be convenient to consider the set $E \subset V \times V$ of (i, j) such that either $i = j$ or $\{i, j\}$ is in \mathcal{E} , rather than \mathcal{E} and, since E and \mathcal{E} carry the same information, to speak about the graph $G = (V, E)$. Any (i, j) such that $i \neq j$ will be called an edge. An r -dimensional Gaussian model is said to be Markov with respect to G if for any edge (i, j) not in E , the i -th and j -th variables are conditionally independent given all the other variables. Such models are known as covariance selection models (see Dempster (1972)) or graphical Gaussian models (see Whittaker (1990) or Lauritzen (1996)). Without loss of generality, we can assume that these models are centered $N_r(0, \Sigma)$ and it is well known that they are characterized by the parameter set P_G of the precision matrices which is the set of positive definite matrices $K = \Sigma^{-1}$ such that $K_{ij} = 0$ whenever the edge (i, j) is not in E . Equivalently, if we denote by M the linear space of symmetric matrices of order r , by $M_r^+ \subset M$ the cone of positive definite (abbreviated > 0) matrices, by I_G the linear space of symmetric incomplete matrices x with missing entries $x_{ij}, (i, j) \notin E$ and by $\pi : M \mapsto I_G$ the projection of M into I_G , the parameter set of the Gaussian model can be described as the set of incomplete matrices $\Sigma_G = \pi(\Sigma)$ with $\Sigma = K^{-1}$ and $K \in P_G$. Indeed it is easy to verify that the entries $\Sigma_{ij}, (i, j) \notin E$ are such that

$$\Sigma_{ij} = \Sigma_{i, V \setminus \{i, j\}} \Sigma_{V \setminus \{i, j\}, V \setminus \{i, j\}}^{-1} \Sigma_{V \setminus \{i, j\}, j},$$

and are therefore not free parameters of the Gaussian models. One can prove that the correspondence between K and the incomplete matrix $\Sigma_G = \pi(\Sigma)$ is one to one. We write $\Sigma_G = \varphi(K) = \pi(K^{-1})$. We note that φ is not explicit when G is not decomposable.

Henceforth in this paper, we will assume that G is decomposable. The reader is referred to Lauritzen (1996) for all the common notions of graphical models used in this paper. We will now simply recall some basic facts and traditional notations, we will use throughout this paper. Every decomposable graph admits a perfect order of its cliques. Let (C_1, \dots, C_k) be such an order. We use the notation $H_1 = R_1 = C_1$ while for $j = 2, \dots, k$ we write

$$H_j = C_1 \cup \dots \cup C_j, \quad R_j = C_j \setminus H_{j-1}, \quad S_j = H_{j-1} \cap C_j.$$

The S_j , $j = 2, \dots, k$ are the minimal separators of G . Some of these separators can be identical. We let $k' \leq k - 1$ denote the number of distinct separators and $\nu(S)$ denote the multiplicity of S that is the number of j such that $S_j = S$. Lauritzen (1996) has proven that the multiplicity $\nu(S)$ of a given minimal separator S is positive and independent of the perfect order of the cliques considered.

For G given decomposable with the set of cliques $\{C_1, \dots, C_k\}$ and $\Sigma^{-1} \in P_G$, the incomplete matrix Σ_G is completely determined by its submatrices $\{\Sigma_{C_i}, i = 1, \dots, k\}$ where, of course, for each $i = 1, \dots, k$, Σ_{C_i} is positive definite. When considering the parameter space of the graphical Gaussian model corresponding to G decomposable, we are therefore led to consider the two cones

$$P_G = \{y \in M_r^+ \mid y_{ij} = 0, (i, j) \notin E\} \quad (1.1)$$

$$Q_G = \{x \in I_G \mid x_{C_i} > 0, i = 1, \dots, k\}. \quad (1.2)$$

Dawid and Lauritzen (1993, Section 7) defined two distributions on Q_G , namely, the hyper Wishart distribution as the distribution of the maximum likelihood estimator of Σ_G , and the hyper inverse Wishart distribution as the Diaconis-Ylvisaker conjugate prior distribution for Σ_G . Subsequently Roverato (2000) gave the distribution of $K = \Sigma^{-1} = \varphi^{-1}(\Sigma_G)$ when Σ_G follows the hyper inverse Wishart distribution. We will call this distribution of K on P_G the G -Wishart. The search for a rich and flexible class of conjugate prior distributions for Σ_G , or equivalently for $K = \Sigma^{-1}$ remains a topic of high interest to statisticians.

When G is complete, $P_G = Q_G = M_r^+$ and we define the regular Wishart distribution on the cone of positive definite matrices of dimension $r = |V|$ by

$$\frac{1}{2^{rp} \Gamma_r(p) |\Sigma|^p} e^{-\frac{1}{2} \text{tr } x \Sigma^{-1}} |x|^p |x|^{-\frac{r+1}{2}} \mathbf{1}_{M_r^+}(x) dx$$

where $p > \frac{r-1}{2}$ is the one-dimensional shape parameter and $\Sigma \in M_r^+$ is the scale parameter.

When G is decomposable, the hyper and hyper inverse Wisharts have been constructed as Markov combination (with respect to G) of the Wishart and its inverse respectively, and so, like the Wishart, they have a one-dimensional shape parameter and a scale parameter in Q_G . Dawid and Lauritzen (1993) have shown that these distributions have Markov properties: the hyper Wishart is weak hyper Markov while the hyper inverse Wishart is strong hyper Markov.

In this paper, we will construct a family of distributions, called type I Wisharts, defined on Q_G and another family, called type II Wisharts, defined on P_G . We shall see in §4 that the inverse of the type II distributions, like the hyper inverse Wishart, form a family of conjugate prior distributions for the scale parameter of the graphical Gaussian model. We

will also show that they are strong directed hyper Markov in the direction given to the graph G by any choice of a perfect numbering of its vertices. This property is parallel to the strong hyper Markov property of the hyper inverse Wishart. We will also show that the type I Wishart is weak hyper Markov, a property parallel to the weak hyper Markov property of the hyper Wishart. The attractive feature of the inverse type II Wishart family of conjugate distributions is that, except in the trivial case where G is complete, the set of shape parameters is of dimension strictly bigger than the number k of cliques in G , thus offering a flexible class of conjugate prior distributions for Σ_G . We shall also note in §4 that it forms a class of enriched conjugate priors for Σ_G in the sense of Consonni and Veronese (2001).

To construct these two families, we define two natural exponential families of distributions affiliated to the Wishart, one on Q_G and one on P_G . Let (C_1, \dots, C_k) denote a perfect order of the cliques of G and let (S_2, \dots, S_k) be its corresponding sequence of minimal separators, some of them being possibly identical. We consider functions of the type

$$H_G(\alpha, \beta; x) = \frac{\prod_{i=1}^k |x_{C_i}|^{\alpha_i}}{\prod_{i=2}^k |x_{S_i}|^{\beta_i}}, \quad x \in Q_G,$$

where α and β are two real-valued functions on the collection \mathcal{C} and \mathcal{S} of cliques and separators respectively such that $\alpha(C_i) = \alpha_i$, $\beta(S_j) = \beta_j$ with $\beta_i = \beta_j$ if $S_i = S_j$. These functions play a very special role in the definition of the two families of distributions that we define. Indeed, if we let $c_i = |C_i|$ and $s_i = |S_i|$ denote the cardinality of C_i and S_i respectively and if we denote

$$\mu_G(dx) = \frac{\prod_{i=1}^k |x_{C_i}|^{-\frac{c_i+1}{2}}}{\prod_{i=2}^k |x_{S_i}|^{-\frac{s_i+1}{2}}} \mathbf{1}_{Q_G}(x) dx,$$

the family of distributions we define on Q_G is, for a given (α, β) , the natural exponential family generated by

$$H_G(\alpha, \beta; x) \mu_G(dx) = H_G\left(\alpha - \frac{1}{2}(c+1), \beta - \frac{1}{2}(s+1); x\right) \mathbf{1}_{Q_G}(x) dx. \quad (1.3)$$

The measure (1.3) can be seen as a Markov combination generalisation of the measure $|x|^p |x|^{-\frac{r+1}{2}} \mathbf{1}_{M_r^+}(x) dx$ generating the Wishart distribution, for a given $p > \frac{r-1}{2}$.

In §3, we will introduce the set \mathcal{A} of (α, β) such that the following integral converges and satisfies

$$\int_{Q_G} e^{-\text{tr}(xy)} H_G(\alpha, \beta; x) \mu_G(dx) = \Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \varphi(y)) \quad (1.4)$$

where $\Gamma_I(\alpha, \beta)$ is some function of (α, β) independent of $y \in P_G$. When (α, β) is in \mathcal{A} we say that $H_G(\alpha, \beta; x)$ has the eigenvalue property with corresponding eigenvalue $\Gamma_I(\alpha, \beta)$

and we define the type I Wishart distribution on Q_G as the distribution with density

$$\frac{1}{\Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \varphi(y))} e^{-\text{tr}(xy)} H_G(\alpha, \beta; x) \mu_G(dx),$$

and with parameters (α, β, y) . In a parallel way, we define a set \mathcal{B} of (α, β) for which an eigenvalue property similar to (1.4) holds for the type II Wishart distribution defined on P_G .

In order to fully describe the type I and II Wishart distributions, it is then necessary to know the sets \mathcal{A} and \mathcal{B} . In §3.2, we show that, for any G , the hyper Wishart and the G -Wishart are particular cases of type I and II distributions, respectively. More precisely, we describe the sets $\mathcal{A}_1 \subset \mathcal{A}$ and $\mathcal{B}_1 \subset \mathcal{B}$ such that for $(\alpha, \beta) \in \mathcal{A}_1$, the type I Wishart is the hyper Wishart and for $(\alpha, \beta) \in \mathcal{B}_1$, the type II Wishart is the G -Wishart. In §3.3, we consider the particular class of decomposable graphs G which do not contain the three-link chain, which we call A_4 , as an induced subgraph. Such graphs are called homogeneous. When G is homogeneous, we describe the sets \mathcal{A} and \mathcal{B} completely and show that they are open sets of dimension $k + k'$, the number of cliques plus the number of distinct separators in G . For G homogeneous, the cones Q_G and P_G are homogeneous and we see that the type I and II Wisharts then belong to the class of Wisharts on homogeneous cones defined by Andersson and Wojnar (2004). In §3.4, we consider nonhomogeneous graphs. In that case, we have, so far, only a partial knowledge of \mathcal{A} and \mathcal{B} . For each perfect order P of the cliques, we define a $(k + 1)$ -dimensional subset A_P of \mathcal{A} such that, for $(\alpha, \beta) \in A_P$, (1.4) holds. We therefore know the subset $\cup A_P$ of \mathcal{A} , but not all of \mathcal{A} . Similarly we define a $(k + 1)$ -dimensional subset B_P of \mathcal{B} such that we know the subset $\cup B_P$ of \mathcal{B} , but not all of \mathcal{B} . We conjecture that the equalities $\mathcal{A} = \cup A_P$ and $\mathcal{B} = \cup B_P$ hold in general for nonhomogeneous graphs and that, thus, the dimension of the manifolds \mathcal{A} and \mathcal{B} is generally $k + 1 < k + k'$. In §3.4, we verify that these two equalities hold when $G = A_4$ by computing, in this case, the 7-uple integral corresponding to (1.4).

In §4, we give the conjugacy and hyper Markov properties mentioned above. We also give the Laplace transforms of the type I and II Wisharts and the expected values of the type I, type II and inverse type II Wisharts. The necessary preliminaries for understanding the cones P_G and Q_G and the measures we define on them are given in §2.1. In §2.2, we give the results needed to work with homogeneous graphs. Most proofs are deferred to the appendix.

2 Preliminaries

2.1 Measures on P_G and Q_G

For the graph $G = (V, E)$, $V = \{1, \dots, r\}$, we write $i \sim j$ to indicate that the edge $\{i, j\}$ is in \mathcal{E} . An undirected graph G is said to be decomposable if it does not contain a cycle of length greater than or equal to four as an induced subgraph and if it is connected. For all the notions related to decomposable graphs that we will introduce below, the reader is referred to Lauritzen (1996, Ch. 2). We denote by Z_G the real linear space of symmetric matrices y of order r such that $y_{ij} = 0$ if $(i, j) \notin E$. We denote by I_G the real linear space of functions $(i, j) \mapsto x_{ij}$ from E to \mathbb{R} such that $x_{ij} = x_{ji}$. The elements of I_G are called *G-incomplete symmetric matrices*. For a decomposable graph, we have defined the cones $P_G \subset Z_G$ and $Q_G \subset I_G$ in (1.1) and (1.2). Recall that M_r^+ denotes the cone of positive definite symmetric matrices of order r . Gröne *et al.* (1984) proved the following.

Proposition 2.1 *When G is decomposable, for any x in Q_G there exists a unique \hat{x} in M_r^+ such that for all (i, j) in E we have $x_{ij} = \hat{x}_{ij}$ and such that \hat{x}^{-1} is in P_G .*

This defines a bijection between P_G and Q_G :

$$\varphi : y = (\hat{x})^{-1} \in P_G \mapsto x = \varphi(y) = \pi(y^{-1}) \in Q_G,$$

where π denotes the projection of M onto I_G . The explicit expression of \hat{x}^{-1} is given in (2.3) below. For $(x, y) \in I_G \times Z_G$, we write $\text{tr}(xy) = \langle x, y \rangle = \sum_{(i,j) \in E} x_{ij}y_{ij}$. By Proposition 2.1, we have for $x \in Q_G$ $\langle x, y \rangle = \text{tr}(\hat{x}y)$, where $\text{tr}(\hat{x}y)$ is defined in the classical way. Thus although xy does not make sense, the notation $\text{tr}(xy)$ is quite convenient. We also use the following notation: if C is a complete subset of vertices and if $x_C = (x_{ij})_{i,j \in C}$ is a matrix, we denote by $(x_C)^0 = (x_{ij})_{i,j \in V}$ the matrix such that $x_{ij} = 0$ for $(i, j) \notin C \times C$.

The following theorem gathers some basic results on decomposable graphs: Part 1 is due to Andersson (private communication), Parts 2 and 3 can be found in Lauritzen (1996, Ch. 5) et Part 4 is due to Roverato (2000).

Theorem 2.1 *Let G be a decomposable graph. Then*

1. *The convex open cones P_G and Q_G are dual to each other in the sense that*

$$P_G = \left\{ y \in Z_G; \text{tr}(xy) > 0 \forall x \in \overline{Q_G} \setminus \{0\} \right\} \quad (2.1)$$

$$Q_G = \left\{ x \in I_G; \text{tr}(xy) > 0 \forall y \in \overline{P_G} \setminus \{0\} \right\} \quad (2.2)$$

2. *For $x \in Q_G$ we have that $y = \hat{x}^{-1}$ is in P_G and*

$$y = \sum_{C \in \mathcal{C}} (x_C^{-1})^0 - \sum_{S \in \mathcal{S}} \nu(S) (x_S^{-1})^0. \quad (2.3)$$

3. For $x \in Q_G$ we have

$$\det \hat{x} = \frac{\prod_{C \in \mathcal{C}} (\det x_C)}{\prod_{S \in \mathcal{S}} (\det x_S)^{\nu(S)}}.$$

4. The absolute value of the Jacobian of the bijection $x \mapsto y = \hat{x}^{-1}$ from Q_G to P_G is

$$\prod_{C \in \mathcal{C}} (\det x_C)^{-|C|-1} \prod_{S \in \mathcal{S}} (\det x_S)^{(|S|+1)\nu(S)}. \quad (2.4)$$

The proof of Part 1 is given in the Appendix. For G complete, Part 4 above becomes the following.

Lemma 2.1 (*Muirhead, 1982*)

The Jacobian of the change of variable $x \in M_r^+ \mapsto y = x^{-1} \in M_r^+$ is $|y|^{-(r+1)}$.

We now introduce the measures which will be the generating measures of the new Wishart exponential families on P_G and Q_G that we are going to define in the next section. Let α and β be two real valued functions on \mathcal{C} and \mathcal{S} respectively. An example of such functions α and β is

$$C \in \mathcal{C} \mapsto \alpha(C) = |C| \quad \text{and} \quad S \in \mathcal{S} \mapsto \beta(S) = |S|.$$

We denote these examples $\alpha = c$ and $\beta = s$. Another example is, for a constant p given,

$$C \in \mathcal{C} \mapsto \alpha(C) = p \quad \text{and} \quad S \in \mathcal{S} \mapsto \beta(S) = p.$$

simply denoted $\alpha = p$ and $\beta = p$. For $x \in Q_G$ we adopt the following notation

$$H_G(\alpha, \beta; x) = \frac{\prod_{C \in \mathcal{C}} (\det x_C)^{\alpha(C)}}{\prod_{S \in \mathcal{S}} (\det x_S)^{\nu(S)\beta(S)}}. \quad (2.5)$$

The functions H_G for the particular case $\alpha = -\frac{1}{2}(c+1)$ and $\beta = -\frac{1}{2}(s+1)$ will play an important role. Indeed, we will use the following as reference measures to generate the exponential families of distributions which are the central object of our study in this paper. These reference measures are

$$\mu_G(dx) = H_G\left(-\frac{1}{2}(c+1), -\frac{1}{2}(s+1); x\right) \mathbf{1}_{Q_G}(x) dx, \quad (2.6)$$

$$\nu_G(dy) = H_G\left(\frac{1}{2}(c+1), \frac{1}{2}(s+1); \varphi(y)\right) \mathbf{1}_{P_G}(y) dy. \quad (2.7)$$

Applying (2.4), we see that ν_G is the image of μ_G by the mapping $x \mapsto y = \hat{x}^{-1}$ and that conversely μ_G is the image of ν_G by $y \mapsto x = \varphi(y)$.

Let M_d^+ denote the cone of positive definite matrices of order d and $L(\mathbb{R}^p, \mathbb{R}^q)$ denote the space of linear transformations from \mathbb{R}^p to \mathbb{R}^q . For $x \in Q_G$, $x_{C_j}, j = 1, \dots, k$ are well

defined and it will be convenient to use the following standard notation for various block submatrices:

$$x_{S_j} = x_{\langle j \rangle}, \quad x_{R_j, S_j} = x_{[j]} = x_{\langle j \rangle}^t, \quad x_{[j]} = x_{R_j}, \quad x_{[j]} = x_{[j]} - x_{[j]} x_{\langle j \rangle}^{-1} x_{\langle j \rangle}, \quad (2.8)$$

where $x_{\langle j \rangle} \in M_{s_j}^+$, $x_{[j]} \in M_{c_j - s_j}^+$, $x_{[j]} \in L(\mathbb{R}^{c_i - s_j}, \mathbb{R}^{s_j})$. It is understood here that $x_{[1]} = x_{[1]} = x_{C_1}$ whereas both $x_{\langle 1 \rangle}$ and $x_{[1]}$ vanish. With this notation, we have, for example, $|\hat{x}| = \prod_{j=1}^k |x_{[j]}|$. In the proof of our main theorems, we will need to split the trace $\langle x, y \rangle$ for $x \in Q_G$ and $y \in P_G$ following a perfect order of the cliques as given in the following lemma.

Lemma 2.2 *Let G be a decomposable graph and let C_1, \dots, C_k be a perfect order of its cliques. For $x \in Q_G$ and $y \in P_G$ with $y = \hat{\sigma}^{-1}$ and $\sigma \in Q_G$, we have*

$$\langle x, y \rangle = \langle x, \hat{\sigma}^{-1} \rangle = \sum_{i=1}^k \left[\langle x_{[i]}, \sigma_{[i]}^{-1} \rangle + \langle (x_{[i]} x_{\langle i \rangle}^{-1} - \sigma_{[i]} \sigma_{\langle i \rangle}^{-1}), \sigma_{[i]}^{-1} (x_{[i]} x_{\langle i \rangle}^{-1} - \sigma_{[i]} \sigma_{\langle i \rangle}^{-1}) x_{\langle i \rangle} \rangle \right]. \quad (2.9)$$

This is a direct consequence of (2.3) and the following standard splitting of the trace for two positive definite matrices $u = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$ and $v = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}$

$$\langle u, v \rangle = \langle u_{11}, v_{11} \rangle + \langle u_{21}, v_{21} \rangle + \langle (u_{21} u_{11}^{-1} + v_{21}^{-1} v_{11}), v_{21} (u_{21} u_{11}^{-1} + v_{21}^{-1} v_{11}) u_{11} \rangle, \quad (2.10)$$

and its corresponding expression if we write $v = \hat{\sigma}^{-1}$ with $\hat{\sigma}$ also positive definite.

$$\langle u, \hat{\sigma}^{-1} \rangle = \langle u_{11}, \sigma_{11}^{-1} \rangle + \langle u_{21}, \sigma_{21}^{-1} \rangle + \langle (u_{21} u_{11}^{-1} - \sigma_{21} \sigma_{11}^{-1}), \sigma_{21}^{-1} (u_{21} u_{11}^{-1} - \sigma_{21} \sigma_{11}^{-1}) u_{11} \rangle. \quad (2.11)$$

We also recall the following basic results that will be used throughout our proofs.

Lemma 2.3 *(Massam & Neher, 1997, Lemma 7) The Jacobian of the change of variables*

$$x \in Q_G \mapsto y = (x_{C_1}, x_{[i]}, x_{[i]} x_{\langle i \rangle}^{-1}, i = 2, \dots, k) \quad (2.12)$$

is

$$\left| \frac{dy}{dx} \right| = \prod_{j=2}^k |x_{\langle j \rangle}|^{c_j - s_j}. \quad (2.13)$$

The following lemma gives a Gaussian distribution we shall use later.

Lemma 2.4 *For x and σ in Q_G , and for $L = L(R^{c_i - s_i}, R^{s_i})$ we have*

$$\int_L e^{-\langle (x_{[i]} x_{\langle i \rangle}^{-1} - \sigma_{[i]} \sigma_{\langle i \rangle}^{-1}), \sigma_{[i]}^{-1} (x_{[i]} x_{\langle i \rangle}^{-1} - \sigma_{[i]} \sigma_{\langle i \rangle}^{-1}) x_{\langle i \rangle} \rangle} d(x_{[i]} x_{\langle i \rangle}^{-1}) = \pi^{\frac{(c_i - s_i) s_i}{2}} \frac{|\sigma_{[i]}|^{-\frac{s_i}{2}}}{|x_{\langle i \rangle}|^{\frac{c_i - s_i}{2}}}. \quad (2.14)$$

The proof follows immediately from Theorem 3.1.1 in Muirhead (1982) by replacing C, D, Y and M in that theorem by $\sigma_{[i]}, x_{<i>}^{-1}, x_{[i>}x_{<i>}^{-1}$ and $\sigma_{[i>}\sigma_{<i>}^{-1}$ respectively.

Let us finally recall the definition of the multivariate Gamma function. For $p > \frac{r-1}{2}$, the r -multivariate Gamma function is

$$\Gamma_r(p) = \pi^{\frac{1}{4}r(r-1)} \prod_{j=1}^r \Gamma(p - \frac{1}{2}(j-1)). \quad (2.15)$$

In the sequel we will need the following two formulas which link multivariate gamma functions of different dimensions. For c and s two positive integers with $s < c$ and for $\alpha > \frac{c-1}{2}$ a real number, we have

$$\pi^{\frac{(c-s)s}{2}} \Gamma_{c-s}(\alpha - \frac{s}{2}) = \frac{\Gamma_c(\alpha)}{\Gamma_s(\alpha)} \quad (2.16)$$

$$\pi^{\frac{(c-s)s}{2}} \Gamma_{c-s}(\alpha) = \frac{\Gamma_c(\alpha)}{\Gamma_s(\alpha - \frac{c-s}{2})} \quad (2.17)$$

2.2 Tools for homogeneous graphs.

In this subsection, we study some properties of homogeneous graphs.

Definition 2.1 *A graph G is said to be homogeneous if it is decomposable and does not contain the graph $\overset{1}{\bullet} - \overset{2}{\bullet} - \overset{3}{\bullet} - \overset{4}{\bullet}$, called A_4 , as an induced subgraph.*

We will see in Theorem 2.2 below why such a graph is called homogeneous. We now need to introduce a number of concepts about undirected graphs.

Definition 2.2 *Given an undirected graph $G = (V, E)$, the associated digraph is the directed graph $G' = (V, E')$ with E' derived from E by the following process. If $i, j \in V$, then the directed edge (i, j) is in E' if and only if*

$$\{i\} \cup nb(i) \supseteq \{j\} \cup nb(j) \quad (2.18)$$

where $nb(i) = \{j; j \neq i, i \sim j\}$.

Note that E' contains all (i, i) for $i \in V$. We write $i \rightarrow j$ if and only if $(i, j) \in E'$. An edge in G can either disappear in G' or become directed or become bi-directed. Note that if $i \sim j$ in G then $i \not\rightarrow j$ if and only if there exists $k, k \neq i, k \neq j$ such that the subgraph of G induced by $\{i, j, k\}$ is $\overset{k}{\bullet} - \overset{j}{\bullet} - \overset{i}{\bullet}$. Note also that if $k \not\sim i$, then it is impossible to have both $i \rightarrow j$ and $k \rightarrow j$. In other words, the configuration $\overset{i}{\bullet} \rightarrow \overset{j}{\bullet} \leftarrow \overset{k}{\bullet}$ in G' is forbidden. Here are two simple examples of digraphs associated to given graphs. For the sake of clarity, the loops $i \rightarrow i$ are not drawn on the digraph G' .

Examples. The graph $\overset{1}{\bullet} - \overset{2}{\bullet}$ becomes the graph $G' \overset{1}{\bullet} \leftrightarrow \overset{2}{\bullet}$. The graph A_4 becomes $\overset{1}{\bullet} \leftarrow \overset{2}{\bullet} \quad \overset{3}{\bullet} \rightarrow \overset{4}{\bullet}$.

It is easy to see from (2.18) that if G is an undirected graph and G' its associated digraph, then the relation $i \rightarrow j$ defined on V is a preorder relation, that is $i \rightarrow j$ and $j \rightarrow k$ implies $i \rightarrow k$. Denote by R the induced equivalence relation defined on V by

$$iRj \Leftrightarrow i \rightarrow j \text{ and } j \rightarrow i \Leftrightarrow \{i\} \cup nb(i) = \{j\} \cup nb(j).$$

Denote by \bar{i} the equivalence class in V/R containing $i \in V$ and denote $\bar{i} \preceq \bar{j}$ the partial order relation on V/R induced by the preorder $i \rightarrow j$. As usual when dealing with partial order, the notation $\bar{i} \prec \bar{j}$ means $\bar{i} \preceq \bar{j}$ and $\bar{i} \neq \bar{j}$. We now introduce the Hasse diagram of V/R .

Definition 2.3 *The Hasse diagram of G is the digraph with vertex set $V_H = V/R$ and with edge set E_H such that an edge (\bar{i}, \bar{j}) is in E_H if*

$$\bar{i} \neq \bar{j}, \bar{i} \preceq \bar{j},$$

and

$$\bar{i} \preceq \bar{k} \preceq \bar{j} \text{ implies either } \bar{k} = \bar{i} \text{ or } \bar{k} = \bar{j}.$$

If $(\bar{i}, \bar{j}) \in E_H$, we write $\bar{i} \rightarrow \bar{j}$. The knowledge of the Hasse diagram of G is equivalent to the knowledge of the partial order relation on V/R . If $\bar{i} \rightarrow \bar{j}$ then \bar{j} is a child of \bar{i} and \bar{i} is a parent of \bar{j} . If i and j are in V it will be convenient to write $i \rightarrow j$ when the corresponding equivalence classes satisfy $\bar{i} \rightarrow \bar{j}$. Let us give an example of construction of a Hasse diagram.

In Figure 1 below, we give a graph G , its associated digraph G' and the corresponding Hasse diagram. In G' , the loops (i, i) are omitted. Since $3 \rightarrow 7$ and $7 \rightarrow 3$, $\{3, 7\}$ is

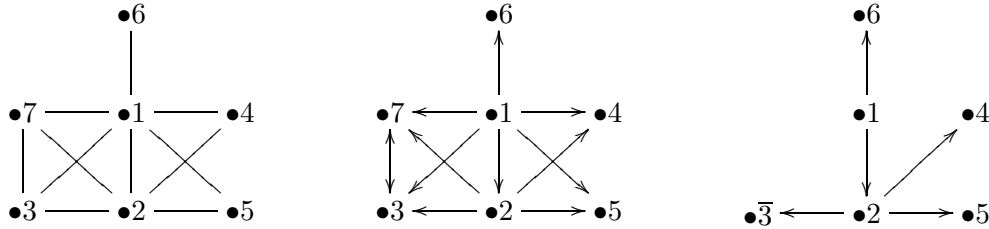


Figure 1: G , G' and the Hasse diagram

an equivalence class denoted by $\bar{3}$ while the 5 other vertices are alone in their equivalence class, which we denote by i rather than \bar{i} for simplicity. On this particular example, the

Hasse diagram is a rooted tree associated to a partial order such that the root 1 is the minimum. The four cliques correspond to the end points of the tree: $C_3 = \{1, 2, 3, 7\}$, $C_4 = \{1, 2, 4\}$, $C_5 = \{1, 2, 5\}$, $C_6 = \{1, 6\}$. The two separators $S_1 = \{1, 2\}$ with multiplicity 2 and $S_2 = \{1\}$ with multiplicity 1 correspond to the other vertices of the diagram. Note that the graph G is homogeneous since it does not contain any A_4 as an induced subgraph. The fact that, in this example, the Hasse diagram is a rooted tree and the graph is homogeneous is not a coincidence since we have the following characterization theorem:

Theorem 2.2 *Let $G = (V, E)$ be a connected graph and let $G' = (V, E')$ be its associated digraph. The following properties are equivalent*

1. G is homogeneous.
2. If $i \sim j$ then either $i \rightarrow j$ or $j \rightarrow i$ in G' .
3. The Hasse diagram of the partially ordered set $(V/R, \preceq)$ is a rooted tree such that its root $\bar{1}$ is the minimal point of V/R and such that the number of children of a vertex is never equal to one.
4. P_G is a homogeneous cone (that is, its automorphism group acts on it transitively).
5. Q_G is a homogeneous cone.

We shall only use equivalences between 1, 2 and 3, which are easy to prove. The equivalence with 4 and 5 is stated for the curiosity of the reader. The homogeneous graphs are specially simple to handle. We call T the set of vertices of the corresponding Hasse tree, so $T = V/R$. Consider the following subset of V

$$V_{\bar{i}} = \cup_{\bar{j} \preceq \bar{i}} \bar{j}.$$

We gather the properties of the Hasse tree of G in the following proposition:

Proposition 2.2 *If T is the Hasse rooted tree of a homogeneous graph G with k cliques and k' minimal separators, we have that*

1. The mapping $\bar{i} \mapsto V_{\bar{i}}$, where $\bar{i} \in T$, gives a one to one correspondence between the cliques and minimal separators of G and, respectively, the end points and non end points of T . In particular, if $k > 1$ the root $\bar{1}$ is a minimal separator which is contained in all minimal separators and cliques of G and the total number of vertices in T is equal to $k + k'$.
2. All orders of the cliques are perfect. The multiplicity $\nu(V_s)$ of a separator V_s is equal to the number of children of s minus one.

Proof: 1. If $\bar{i} \in T$ then we observe that $V_{\bar{i}}$ is complete since if j and l are in $V_{\bar{i}}$ then either $\bar{j} \preceq \bar{l} \preceq \bar{i}$ or $\bar{l} \preceq \bar{j} \preceq \bar{i}$. In both cases $j \sim l$. Conversely, if $C \subset V$ is complete then $\bar{C} = \cup\{\bar{j} \in T ; j \in C\}$ is contained in some $V_{\bar{i}}$. If not there exist \bar{j} and \bar{l} in \bar{C} which are not comparable in the poset T and therefore $j \not\sim l$ which contradicts the fact that C is complete. Thus the maximal cliques are the $V_{\bar{i}}$'s where $\bar{i} \in T$ has no children, that is \bar{i} is an end point. Finally, if $\bar{i} \in T$ has children \bar{j} and \bar{l} then $V_{\bar{i}}$ is a minimal separator of j and l as can easily be seen. Conversely, if j and l are in V with $j \not\sim l$ there exists a unique minimal separator between them which is $V_{\bar{i}}$ where $\bar{i} = \max\{s \in T ; s \preceq \bar{j}, s \preceq \bar{l}\}$.

2. Consider any order $(t(1), \dots, t(k))$ of the endpoints of the tree and the corresponding order $(V_{t(1)}, \dots, V_{t(k)})$ of the cliques. For $j = 2, \dots, k$ and for $l = 1, \dots, j - 1$

$$s(l) = \max\{s \in T ; s \prec t(j), s \prec t(l)\}.$$

Since $\bar{1} \preceq s(l) \prec t(j)$ for all $l = 1, \dots, j - 1$ then $s(l_j) = \max\{s(l) ; l = 1, \dots, j - 1\}$ exists and

$$V_{s(l_j)} = \left(V_{t(1)} \cup \dots \cup V_{t(j-1)} \right) \cap V_{t(j)}$$

is a minimal separator contained in the clique $V_{t(l_j)}$ with $l_j < j$. Thus the order is perfect.

Now, given a minimal separator V_s , we show that the number $\nu(V_s)$ of j such that there exists l_j with $1 \leq l_j < j \leq k$ and $s = s(l_j)$, where $(l_j, s(l_j))$ is as defined above, is equal to $c(s) - 1$ where $c = c(s)$ is the number of children of s . Suppose first that $\nu(V_s) \geq c$. Then there exists endpoints $t(j_1), \dots, t(j_c)$ of T such that $j_1 < \dots < j_c$ and such that $s = s(l_{j_1}) = \dots = s(l_{j_c})$. Thus $s \prec t(j_1), \dots, s \prec t(j_c)$. Furthermore $l_{j_1} < j_1$ and $s \prec t(l_{j_1})$. This implies that s has at least $c + 1$ children, a contradiction. Thus $\nu(V_s) \leq c(s) - 1$. Finally, one sees by induction that the number of edges of an undirected tree is the number of vertices minus one. Since $\sum_s c(s)$ is equal to the number of edges in the graph, this implies that $\sum_s (c(s) - 1) = k - 1$, where the sum is taken over the non end points s of T . To conclude the proof, we use the fact that by definition of the multiplicity of a minimal separator, the sum of the $\nu(V_s)$ is also $k - 1$. Thus $\sum_s [c(s) - 1 - \nu(V_s)] = (k - 1) - (k - 1) = 0$. Since we have a null sum of non negative terms we get $\nu(V_s) = c(s) - 1$ for all minimal separators. \square

It follows from the proposition above that there is a one to one correspondence between the set of homogeneous graphs and the set of rooted trees with vertices weighted by positive integers and such that no vertex has exactly one child. Note that a complete graph is homogeneous. It is characterized by the fact that its Hasse diagram is just a point. A decomposable graph with only one separator is homogeneous. Its Hasse tree looks like a daisy. An undirected tree is decomposable but is not homogeneous in general. Finally it

is possible to prove that if all orders of the cliques of a decomposable graph G are perfect then G is homogeneous.

3 The Wishart families of type I and II

In this section, we define two families of Wishart distributions. We will study special cases in §3.2, the homogeneous case in §3.3 and the non homogeneous case in §3.4.

3.1 Definitions

Consider the two integrals

$$I(\alpha, \beta; y) = \int_{Q_G} e^{-\langle x, y \rangle} H_G(\alpha, \beta; x) \mu_G(dx) \quad \text{for } y \in P_G \quad (3.1)$$

$$J(\alpha, \beta; x) = \int_{P_G} e^{-\langle x, y \rangle} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) \quad \text{for } x \in Q_G. \quad (3.2)$$

We define \mathcal{A} to be the set of (α, β) such that $I(\alpha, \beta; y)$ converges for all $y \in P_G$ and such that $y \mapsto \frac{I(\alpha, \beta; y)}{H_G(\alpha, \beta; \varphi(y))}$ is a constant on P_G . This constant is a function on \mathcal{A} that we denote by $\Gamma_I(\alpha, \beta)$. Similarly we define \mathcal{B} to be the set of (α, β) such that $J(\alpha, \beta; x)$ converges for all $x \in Q_G$ and such that $x \mapsto \frac{J(\alpha, \beta; x)}{H_G(\alpha, \beta; x)}$ is a constant on Q_G . This constant is a function on \mathcal{B} that we denote by $\Gamma_{II}(\alpha, \beta)$. The sets \mathcal{A} and \mathcal{B} will be studied in §3.3, §3.4 and §3.5.

We note here that since $\mu_G(du)$ is the image of the measure $\nu_G(dy)$ under the mapping $y \mapsto u = \varphi(y)$ (see §2.1), (3.2) can be written

$$J(\alpha, \beta; x) = \int_{Q_G} e^{-\langle x, \hat{u}^{-1} \rangle} H_G(\alpha, \beta; u) \mu_G(du) \quad \text{for } x \in Q_G.$$

This expression of (3.2) and the passage from $y \in P_G$ to $u = \varphi(y) \in Q_G$ will be used several times in the remainder of the paper for defining the inverse type II Wishart and to perform various computations. The Wishart distributions of type I will be the probabilities

$$\frac{1}{\Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \varphi(y))} e^{-\langle x, y \rangle} H_G(\alpha, \beta; x) \mu_G(dx),$$

defined on Q_G and indexed by the parameters $(\alpha, \beta; y)$ in $\mathcal{A} \times P_G$. To follow the standard notation for distributions related to the Wishart, when $y \in P_G$ is the parameter of the type I Wishart, we often write $y = \hat{\sigma}^{-1}$ with $\sigma \in Q_G$ so that, for $\sigma \in Q_G$, $(\alpha, \beta) \in \mathcal{A}$, the type I Wishart distribution can be written

$$W_{Q_G}(\alpha, \beta, \sigma; dx) = e^{-\langle x, \hat{\sigma}^{-1} \rangle} \frac{H_G(\alpha, \beta; x)}{\Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \sigma)} \mu_G(dx). \quad (3.3)$$

The Wishart distributions of type II will be the probabilities

$$W_{P_G}(\alpha, \beta, \theta; dy) = e^{-\langle \theta, y \rangle} \frac{H_G(\alpha, \beta; \varphi(y))}{\Gamma_{II}(\alpha, \beta) H_G(\alpha, \beta; \theta)} \nu_G(dy). \quad (3.4)$$

defined on P_G and indexed by the parameters $(\alpha, \beta; \theta)$ in $\mathcal{B} \times Q_G$. We therefore consider the following two natural exponential families.

Definition 3.1 For $(\alpha, \beta) \in \mathcal{A}$, the type I Wishart family of distributions is defined by

$$\mathcal{F}_{(\alpha, \beta), I} = \{W_{Q_G}(\alpha, \beta, \sigma; dx), \sigma \in Q_G\}. \quad (3.5)$$

Definition 3.2 For $(\alpha, \beta) \in \mathcal{B}$, the type II Wishart family of distributions is defined by

$$\mathcal{F}_{(\alpha, \beta), II} = \{W_{P_G}(\alpha, \beta, \theta; dy), \theta \in Q_G\}. \quad (3.6)$$

Following the pattern of what is done for the Wishart distribution, we now define Type I and II inverse Wishart and F distributions.

Definition 3.3 Let G be given. If $X \sim W_{Q_G}(\alpha, \beta, \sigma)$ where $(\alpha, \beta) \in \mathcal{A}$ and $\sigma \in Q_G$, then $Y = \hat{X}^{-1}$ is said to follow the inverse type I Wishart, defined on P_G , and its distribution is

$$IW_{Q_G}(\alpha, \beta, \sigma; dy) = \frac{e^{-\langle \varphi(y), \hat{\sigma}^{-1} \rangle} H_G(\alpha, \beta; \varphi(y))}{\Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \sigma)} \nu_G(dy) \quad (3.7)$$

The distribution (3.7) is clearly immediately derived from the distribution (3.3) by recalling that $x = \varphi(y)$ and that $\nu_G(dy)$ is the image of $\mu_G(dx)$ by the mapping $x \mapsto y = \hat{x}^{-1}$.

Definition 3.4 Let G be given. If $Y \sim W_{P_G}(\alpha, \beta, \theta)$ where $(\alpha, \beta) \in \mathcal{B}$ and $\theta \in Q_G$, then $X = \varphi(Y)$ is said to follow the inverse type II Wishart, defined on Q_G , and its distribution is

$$IW_{P_G}(\alpha, \beta, \theta; dx) = \frac{e^{-\langle \theta, \hat{x}^{-1} \rangle} H_G(\alpha, \beta; x)}{\Gamma_{II}(\alpha, \beta) H_G(\alpha, \beta; \theta)} \mu_G(dx) \quad (3.8)$$

Here too, density (3.8) is immediately derived from (3.4).

Let $\mathcal{B} - \mathcal{A} = \{(\alpha' - \alpha, \beta' - \beta) : (\alpha', \beta') \in \mathcal{B}, (\alpha, \beta) \in \mathcal{A}\}$. Since $\mathcal{B} - \mathcal{A} \subset \mathcal{B}$ and $\mathcal{A} - \mathcal{B} \subset \mathcal{A}$ are false in general as will be seen, for example, when $G = A_4$, to give the following definition of the F distributions, we will have to insure that the parameters $\alpha' - \alpha$ and $\beta' - \beta$ are in the correct sets.

Definition 3.5 Let θ and σ be in P_G and Q_G , and let $(\alpha, \beta) \in \mathcal{A}$ and $(\alpha', \beta') \in \mathcal{B}$. Then

1. for $(\alpha' - \alpha, \beta' - \beta) \in \mathcal{B}$, $\sigma \in Q_G$, the F distribution of the first kind with parameters $(\alpha, \beta, \alpha', \beta', \sigma)$ is the following distribution on Q_G

$$\frac{\Gamma_{II}(\alpha' - \alpha, \beta' - \beta)}{\Gamma_I(\alpha, \beta) \Gamma_{II}(\alpha', \beta')} H_G(-\alpha', -\beta'; \sigma) H_G(\alpha' - \alpha, \beta' - \beta; \sigma + x) H_G(\alpha, \beta; x) \mu_G(dx);$$

2. for $(\alpha - \alpha', \beta - \beta') \in \mathcal{A}$, $\theta \in Q_G$, the F distribution of the second kind with parameters $(\alpha, \beta, \alpha', \beta', \theta)$ is the following distribution on P_G

$$\frac{\Gamma_I(\alpha - \alpha', \beta - \beta')}{\Gamma_I(\alpha, \beta) \Gamma_{II}(\alpha', \beta')} H_G(-\alpha, -\beta; \varphi(\theta)) H_G(\alpha - \alpha', \beta - \beta'; \varphi(\theta + y)) H_G(\alpha', \beta'; \varphi(y)) \nu_G(dy).$$

Note here again that the lack of multiplicative structure on P_G and Q_G when G is not complete prevents us from relating these distributions to some form of quotient X/X' of independent random variables with distributions $W_{Q_G}(\alpha, \beta; \sigma)$ and $W_{Q_G}(\alpha', \beta'; \sigma)$, respectively. A study of the multivariate F distribution when G is complete can be found in Olkin & Rubin (1964). We could also define rather explicitly Beta distributions of type I by introducing the conditional distributions $X|X + X'$ where $X \sim W_{Q_G}(\alpha, \beta, \sigma)$ and $X' \sim W_{Q_G}(\alpha', \beta', \sigma)$ are independent such that $(\alpha + \alpha', \beta + \beta')$ is still in \mathcal{A} . Again, since the cone Q_G has no special multiplicative structure, these beta distributions unfortunately do not seem to enjoy properties linking them to some ratio analogous to $X/(X + X')$ as it happens when the graph is complete. The same problem arises with Beta distributions of type II. Finally, we could also consider the distribution

$$\frac{e^{-\langle x, y \rangle} H_G(\alpha, \beta; x)}{I(\alpha, \beta; y)} \mu_G(dx)$$

where we only require that (α, β) be such that $I(\alpha, \beta; y)$ defined by (3.1) converges. Under such generality, these distributions have no interesting properties: their Laplace transforms are not explicit, their family is not stable by convolution as our Wishart distributions of type I are (see Proposition 3.2) and they have no hyper Markov property. A similar remark holds for J defined by (3.2) and type II Wisharts.

3.2 The hyper and inverse hyper inverse Wishart distributions

We first observe that when G is complete, both type I and type II Wishart distributions coincide with the ordinary Wishart distribution. We will see now that for special values of (α, β) , the type I and II Wisharts are, respectively, the hyper Wishart as defined by Dawid and Lauritzen (1993) and the G -Wishart first identified by Roverato (2000) as the inverse of the hyper inverse Wishart defined also by Dawid and Lauritzen (1993). To describe these distributions, it is convenient to fix a perfect order of the cliques.

The hyper Wishart on Q_G . Let G be given and let p be a scalar. Let \mathcal{A}_1 be the one-dimensional subset of $\mathbb{R}^{k+k'}$ defined as

$$\mathcal{A}_1 = \{(\alpha, \beta) \mid \alpha(C) = p, C \in \mathcal{C}, \beta(S) = p, S \in \mathcal{S} \text{ with } p > \max_{C \in \mathcal{C}} \frac{1}{2}(|C| - 1)\},$$

For $(\alpha, \beta) \in \mathcal{A}_1$ we then have

$$W_{Q_G}(\alpha, \beta, \sigma; dx) \propto \frac{\prod_{i=1}^k w_{c_i}(p, \sigma_{C_i}; x_{C_i})}{\prod_{i=2}^k w_{s_i}(p, \sigma_{S_i}; x_{S_i})} \mathbf{1}_{Q_G}(x) dx \quad (3.9)$$

with

$$w_{c_i}(p, \sigma_{C_i}; x_{C_i}) = \frac{|x_{C_i}|^{p - \frac{c_i + 1}{2}}}{\Gamma_{c_i}(p) |\sigma_{C_i}|^p} e^{-\langle x_{C_i}, \sigma_{C_i}^{-1} x_{C_i} \rangle}.$$

We note that the expression of $W_{Q_G}(\alpha, \beta, \sigma; dx)$ in (3.9) does not depend on the chosen perfect order of the cliques. The expression in the right-hand side of (3.9) is a Markov combination of Wishart distributions with shape parameter p and scale parameter σ_{C_i} and σ_{S_i} on the cliques and separators of G respectively. By Theorem 2.6 of Dawid and Lauritzen (1993), it is a distribution. It is in fact the hyper Wishart distribution, as defined in that same paper. Therefore both sides of (3.9) are equal and equal to the density of the hyper Wishart distribution and it follows immediately that $\mathcal{A}_1 \subset \mathcal{A}$ for any given G and

$$\Gamma_I(\alpha, \beta) = \frac{\prod_{i=1}^k \Gamma_{c_i}(p)}{\prod_{i=2}^k \Gamma_{s_i}(p)}.$$

The G -Wishart on P_G . Let G be given and let $\delta > 0$ be a scalar.

Let \mathcal{B}_1 be the one-dimensional subset of $\mathbb{R}^{k+k'}$ defined as

$$\mathcal{B}_1 = \{(\alpha, \beta) \mid \alpha(C) = -\frac{1}{2}(\delta + |C| - 1), C \in \mathcal{C}, \beta(S) = -\frac{1}{2}(\delta + |S| - 1), S \in \mathcal{S}, \delta > 0\}.$$

For $(\alpha, \beta) \in \mathcal{B}_1$ and for $x = \varphi(y)$

$$\begin{aligned} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) &= \frac{\prod_{i=1}^k |x_{C_i}|^{-\frac{\delta+c_i-1}{2} + \frac{c_i+1}{2}}}{\prod_{i=2}^k |x_{S_i}|^{-\frac{\delta+s_i-1}{2} + \frac{s_i+1}{2}}} \mathbf{1}_{P_G}(y) dy \\ &= \frac{\prod_{i=1}^k |x_{C_i}|^{-\frac{\delta-2}{2}}}{\prod_{i=2}^k |x_{S_i}|^{-\frac{\delta-2}{2}}} \mathbf{1}_{P_G}(y) dy = |y|^{\frac{\delta-2}{2}} \mathbf{1}_{P_G}(y) dy, \end{aligned}$$

where, as before, the expression of $H_G(\alpha, \beta; \varphi(y)) \nu_G(dy)$ does not depend on any chosen perfect order of the cliques. Therefore, $W_{P_G}(\alpha, \beta, \theta; dy) \propto |y|^{\frac{\delta-2}{2}} e^{-\langle \theta, y \rangle} dy$ is the G -Wishart distribution first identified by Roverato (2000) as the inverse of the hyper inverse Wishart. It follows immediately that $\mathcal{B}_1 \subset \mathcal{B}$, that

$$\Gamma_{II}(\alpha, \beta) = \frac{\prod_{i=1}^k \Gamma_{c_i}\left(\frac{\delta+c_i-1}{2}\right)}{\prod_{i=2}^k \Gamma_{s_i}\left(\frac{\delta+s_i-1}{2}\right)}$$

and that the type II Wishart is the G -Wishart defined on P_G for $\delta > 0$, $\theta \in Q_G$. The distribution of this special type II Wishart is

$$W_{P_G}(\alpha, \beta, \theta) = \frac{\prod_{i=1}^k |\theta_{C_i}|^{\frac{\delta+c_i-1}{2}} \prod_{i=2}^k \Gamma_{s_i}\left(\frac{\delta+s_i-1}{2}\right)}{\prod_{i=1}^k \Gamma_{c_i}\left(\frac{\delta+c_i-1}{2}\right) \prod_{i=2}^k |\theta_{S_i}|^{\frac{\delta+s_i-1}{2}}} |y|^{\frac{\delta-2}{2}} e^{-\langle \theta, y \rangle} \mathbf{1}_{P_G}(y) dy.$$

3.3 The homogeneous case

We now consider the type I and II Wishart distributions when the graph G is homogeneous as defined in §2.2. Our aim is to identify the sets \mathcal{A} and \mathcal{B} and the value of the normalising

constants $\Gamma_I(\alpha, \beta)$ and $\Gamma_{II}(\alpha, \beta)$. It is convenient to introduce the following notation, consistent with the notation introduced in (2.8). For $u \in V/R$, we define

$$x_{[u]} = (x_{ij})_{\bar{i}=\bar{j}=u}, \quad x_{[u>} = (x_{ij})_{\bar{i}=u, \bar{j}<u}, \quad x_{<u>} = (x_{ij})_{\bar{i}<u, \bar{j}<u}, \quad x_{[u] \cdot} = x_{[u]} - x_{[u>} x_{<u>}^{-1} x_{<u]} \quad .$$

We also adopt the convention that a vertex of the Hasse tree (T, E_H) will be denoted by t if it is an end point of the tree and by q if it is not an end point. From Proposition 2.2, to each t corresponds a unique clique $C_t = \cup_{u \in T, u \preceq t} [u]$ and therefore a number $\alpha_t = \alpha(C_t)$. And, to each q corresponds a unique minimal separator $S_q = \cup_{u \in T, u \preceq q} [u]$ and therefore a number $\beta_q = \beta(S_q)$. The positive integer $\nu(q)$ is the number of children of q minus one. From Proposition 2.2, this is also the multiplicity of S_q . With these conventions, for each $u \in T$, we write

$$\rho_u = \rho_u(\alpha, \beta) = \sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q) \beta_q.$$

We define n_u to be the cardinality of the vertex u of the Hasse tree, that is the number of vertices in V that are in the vertex u of the Hasse tree of G . We also define

$$m_u = n_u + \sum_{v \prec u} n_v = \sum_{v \preceq u} n_v. \quad (3.10)$$

The following two theorems give \mathcal{A} , \mathcal{B} and the corresponding normalising constants for the Wisharts of type I and II in the homogeneous case.

Theorem 3.1 *Let G be a homogeneous graph. Then*

$$\mathcal{A} = \{(\alpha, \beta) \mid \rho_u > \frac{1}{2}(\sum_{v \preceq u} n_v - 1), u \in T\}.$$

More specifically for $(\alpha, \beta) \in \mathcal{A}$ and $\sigma \in Q_G$, the integral (3.1) converges and

$$\begin{aligned} \int_{Q_G} e^{-\langle x, \hat{\sigma}^{-1} \rangle} H_G(\alpha, \beta; x) \mu_G(dx) &= \prod_{u \in T} \pi^{\sum_{v \prec u} \frac{n_u n_v}{2}} |\sigma_{[u]}|^{\rho_u} \Gamma_{n_u}(\rho_u - \sum_{v \prec u} \frac{n_v}{2}) \quad (3.11) \\ &= H_G(\alpha, \beta; \sigma) \prod_{u \in T} \pi^{\sum_{v \prec u} \frac{n_u n_v}{2}} \Gamma_{n_u}(\rho_u - \sum_{v \prec u} \frac{n_v}{2}) \end{aligned}$$

Theorem 3.2 *Let G be a homogeneous graph. Then*

$$\mathcal{B} = \{(\alpha, \beta) \mid -\rho_u > \frac{1}{2}(\sum_{v \succeq u} n_v - 1), u \in T\}.$$

More specifically for $(\alpha, \beta) \in \mathcal{B}$ and $\theta \in Q_G$, the integral (3.2) converges and

$$\begin{aligned} \int_{P_G} e^{-\langle y, \theta \rangle} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) &= \prod_{u \in T} \pi^{\sum_{v \prec u} \frac{n_u n_v}{2}} |\theta_{[u]}|^{-\rho_u} \Gamma_{n_u}(-\rho_u - \sum_{v \succ u} \frac{n_v}{2}) \quad (3.12) \\ &= H_G(\alpha, \beta; \theta) \prod_{u \in T} \pi^{\sum_{v \prec u} \frac{n_u n_v}{2}} \Gamma_{n_u}(-\rho_u - \sum_{v \succ u} \frac{n_v}{2}). \end{aligned}$$

We note that the parameter sets \mathcal{A} and \mathcal{B} in the homogeneous case are $k + k'$ -dimensional. The proof of Theorem 3.1 follows the same line as that of the proof of Theorem 3.3 given in the Appendix. It is based on Proposition 3.1 given below and on the following analog of formula (2.9) for the traces: for x and σ in Q_G ,

$$\langle x, \hat{\sigma}^{-1} \rangle = \sum_{u \in T} \left[\langle x_{[u] \cdot}, \sigma_{[u] \cdot}^{-1} \rangle + \langle (x_{[u] >} x_{<u>}^{-1} - \sigma_{[u] >} \sigma_{<u>}^{-1}), \sigma_{[u] \cdot}^{-1} (x_{[u] >} x_{<u>}^{-1} - \sigma_{[u] >} \sigma_{<u>}^{-1}) x_{<u>} \rangle \right] \quad (3.13)$$

where it is understood that, as in (2.8) and (2.9), for $u = \bar{1}$, the root of T , the summand reduces to $\langle x_1, \sigma_1^{-1} \rangle$. Then, using Proposition 3.1 and formula (3.13), the integral in (3.11) is obtained by a series of standard integrations. The proof of (3.13) is parallel to the proof of (2.9) and will not be given here. The proof of Proposition 3.1 is given in the Appendix.

Proposition 3.1 *For G homogeneous, the image of $H_G(\alpha, \beta; x) \mu_G(dx)$ under the mapping*

$$x = (x_{[u]}, x_{[u] >}, u \in T) \mapsto (x_{[u]}, x_{[u] >} x_{<u>}^{-1}, u \in T) \quad (3.14)$$

is

$$H_G^*(\alpha, \beta; x) \mu_G^*(dx_{[u] \cdot}, dx_{[u] >} x_{<u>}^{-1}, u \in T) = \prod_{u \in T} |x_{[u] \cdot}|^{\lambda_u - \frac{n_u + 1}{2}} dx_{[u] \cdot} d(x_{[u] >} x_{<u>}^{-1}),$$

where

$$\lambda_u = \rho_u + \sum_{v > u} \frac{n_v}{2} - \sum_{v < u} \frac{n_v}{2}.$$

The proof of Theorem 3.2 also follows the general lines of the proof of Theorem 3.4 given in the appendix. We first observe that the image of $H_G(\alpha, \beta; \varphi(y)) \nu_G(dy)$ under the change of variable $y \mapsto x = \varphi(y)$ is $H_G(\alpha, \beta; x) \mu_G(dx)$ so that

$$\int_{P_G} e^{-\langle y, \theta \rangle} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) = \int_{Q_G} e^{-\langle \theta, \hat{x}^{-1} \rangle} \frac{\prod_{j=1}^k |x_{C_j}|^{\alpha_j - \frac{c_j + 1}{2}}}{\prod_{j=2}^k |x_{S_j}|^{\beta_j - \frac{s_j + 1}{2}}} dx,$$

where, as usual, the integral on the right hand side of the equation above does not depend upon the chosen perfect order of the cliques. We then use (3.14) and (3.13) applied to $\langle \theta, \hat{x}^{-1} \rangle$ to obtain the expression of the integral in (3.12) by a series of standard integrations.

Using Proposition 3.1, (3.13) and (3.11) it is fairly straightforward to show that the image of the type I Wishart by the change of variable (3.14) is the following distribution

$$\begin{aligned} & W_{Q_G}^*(\alpha, \beta, \sigma; dx_{[u] \cdot}, d(x_{[u] >} x_{<u>}^{-1}), u \in T) \quad (3.15) \\ &= \prod_{u \in T} \left[\frac{|x_{[u] \cdot}|^{\lambda_u - \frac{n_u + 1}{2}} e^{-\langle x_{[u] \cdot}, \sigma_{[u] \cdot}^{-1} \rangle} e^{-\langle (x_{[u] >} x_{<u>}^{-1} - \sigma_{[u] >} \sigma_{<u>}^{-1}), \sigma_{[u] \cdot}^{-1} (x_{[u] >} x_{<u>}^{-1} - \sigma_{[u] >} \sigma_{<u>}^{-1}) x_{<u>} \rangle}}{\pi^{\frac{1}{2} n_u (\sum_{v > u} n_v)} |\sigma_{[u] \cdot}|^{\rho_u} \Gamma_{n_u}(\lambda_u - \frac{1}{2} \sum_{v < u} n_v)} \right. \\ & \quad \left. \mathbf{1}_{D_u}(x_{[u] \cdot}, (x_{[u] >} x_{<u>}^{-1})) dx_{[u] \cdot} d(x_{[u] >} x_{<u>}^{-1}) \right]. \end{aligned}$$

where $D_u = (M_{n_u}^+ \times L(\mathbb{R}^{n_u}, \mathbb{R}^{m_u - n_u}))$. This distribution is exactly the distribution of the Wisharts defined by Andersson and Wojnar (2004) on homogeneous cones. Therefore when G is a homogeneous graph the type I Wisharts coincide with the Wisharts of Andersson and Wojnar (2004) for the homogeneous cone Q_G corresponding to G . Since the dual P_G of Q_G is also homogeneous, we could also show that the type II Wisharts correspond to the Wisharts as defined by Andersson and Wojnar on P_G . However, there are many other homogeneous cones not of the form P_G and Q_G . Our calculations are simpler and self contained in the particular cases that we investigate here.

Using (3.15), Proposition 3.1, (3.13) and (3.12), we obtain the image of the inverse type II Wishart by the change of variable (3.14). The image of the distribution of $X \sim IW_{P_G}(\alpha, \beta, \theta)$, the inverse of the type II Wishart when G is homogeneous is given by

$$\begin{aligned}
& IW_{P_G}^*(\alpha, \beta, \theta; dx_{[u]}, dx_{[u>}x_{<u>}^{-1}, u \in T) \tag{3.16} \\
&= \prod_{u \in T} \left[\frac{|x_{[u]}|^{\lambda_u - \frac{n_u + 1}{2}} e^{-\langle x_{[u]}^{-1}, \theta_{[u]}^{-1} \rangle} e^{-\langle (x_{[u]}x_{<u>}^{-1} - \theta_{[u>} \theta_{<u>}^{-1}), \theta_{<u>} \rangle} (x_{<u>}^{-1} x_{<u>} - \theta_{<u>}^{-1} \theta_{<u>}) x_{[u]}^{-1} \rangle}{\pi^{\frac{1}{2}n_u} \sum_{v \prec u} n_v |\theta_{[u]}|^{-\rho_u} \Gamma_{n_u}(-\rho_u - \sum_{v \succ u} \frac{n_v}{2})} \right. \\
& \quad \left. \mathbf{1}_{D_u}(x_{[u]}, (x_{[u>}x_{<u>}^{-1})) dx_{[u]} d(x_{[u>}x_{<u>}^{-1}) \right].
\end{aligned}$$

Example: Consider the following graph G_0 .

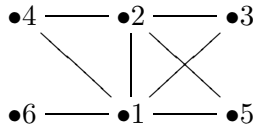


Figure 2: Graph G_0

We index each clique according to the vertex of the Hasse tree of G_0 which represents it. Thus $C_3 = \{1, 2, 3\}$, $C_4 = \{1, 2, 4\}$, $C_5 = \{1, 2, 5\}$, $C_6 = \{1, 6\}$. Minimal separators are $S_1 = \{1\}$ and $S_2 = \{1, 2\}$ with $\nu(S_2) = 2$ and $\nu(S_1) = 1$. We set

$$\alpha(C_3) = \alpha_3, \alpha(C_4) = \alpha_4, \alpha(C_5) = \alpha_5, \alpha(C_6) = \alpha_6, \beta(S_1) = \beta_1, \beta(S_2) = \beta_2.$$

The Hasse tree corresponding to G_0 is identical to the Hasse diagram of Figure 1, with $\bar{3}$ replaced by 3. Since the cardinality of all the vertices of the Hasse tree is 1, for the sake of simplicity we will denote the vertices of the tree by 1, 2, 3, 4, 5 and 6 so that we have

$$n_i = 1, i = 1, \dots, 6 \text{ and } 1 \preceq 6, 1 \preceq 2, 2 \preceq 3, 2 \preceq 4, 2 \preceq 5,$$

$\rho_1 = \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 - \beta_1 - 2\beta_2$, $\rho_2 = \alpha_3 + \alpha_4 + \alpha_5 - 2\beta_2$, $\rho_3 = \alpha_3$, $\rho_4 = \alpha_4$, $\rho_5 = \alpha_5$, $\rho_6 = \alpha_6$,
and $\lambda_1 = \rho_1 + \frac{5}{2}$, $\lambda_2 = \rho_2 + \frac{3}{2} - \frac{1}{2}$, $\lambda_3 = \rho_3 - \frac{2}{2}$, $\lambda_4 = \rho_4 - \frac{2}{2}$, $\lambda_5 = \rho_5 - \frac{2}{2}$, $\lambda_6 = \rho_6 - \frac{1}{2}$.

Therefore,

$$\begin{aligned} \mathcal{A} &= \{(\alpha, \beta) \mid \rho_1 > 0, \rho_2 > \frac{1}{2}, \rho_i > 1, i = 3, 4, 5, \rho_6 > \frac{1}{2}\} \\ &= \{(\alpha, \beta) \mid \alpha_i > 1, i = 3, 4, 5, \alpha_6 > \frac{1}{2}, \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 - 2\beta_2 - \beta_1 > 0, \\ &\hspace{15em} \alpha_3 + \alpha_4 + \alpha_5 - 2\beta_2 > \frac{1}{2}\} \end{aligned} \quad (3.17)$$

$$\begin{aligned} \mathcal{B} &= \{(\alpha, \beta) \mid \rho_1 < -\frac{5}{2}, \rho_2 < -\frac{3}{2}, \rho_i < 0, i = 3, 4, 5, 6\} \\ &= \{(\alpha, \beta) \mid \alpha_i < 0, i = 3, 4, 5, 6, \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 - 2\beta_2 - \beta_1 < -\frac{5}{2}, \\ &\hspace{15em} \alpha_3 + \alpha_4 + \alpha_5 - 2\beta_2 < -\frac{3}{2}\} \end{aligned} \quad (3.18)$$

3.4 The nonhomogeneous case

We now consider a nonhomogeneous graph G , i.e. a graph containing A_4 as an induced subgraph. As in the case of homogeneous graphs, our aim is to identify \mathcal{A}, \mathcal{B} and the corresponding eigenvalues. We will see that we are, in fact, only able to identify a subset of \mathcal{A} and \mathcal{B} and the corresponding eigenvalues $\Gamma(\alpha, \beta)$. The results are given in Theorem 3.3 and Theorem 3.4 below. For G a noncomplete decomposable graph, let $P = (C_1, \dots, C_k)$ be a perfect order of the family \mathcal{C} of its cliques and (S_2, \dots, S_k) be the associated sequence of minimal separators. Recall that $c_j = |C_j|$ and $s_j = |S_j|$ denote the cardinality of C_j and S_j respectively. For given α and β we write $\alpha_j = \alpha(C_j)$ and $\beta_j = \beta(S_j)$. For a given minimal separator S we write

$$J(P, S) = \{j = 2, \dots, k \mid S_j = S\},$$

and for a given perfect order P of the cliques, we define A_P to be the set of (α, β) such that

1. $\sum_{j \in J(P, S)} \alpha_j - \nu(S)\beta(S) = 0$, for all S different of S_2 ;
2. $\alpha_j - \frac{c_j - 1}{2} > 0$ for all $C_j \in \mathcal{C}$;
3. $\alpha_1 + \delta_2 > \frac{s_2 - 1}{2}$ where $\delta_2 = \sum_{j \in J(P, S_2)} \alpha_j - \nu(S_2)\beta_2$.

Recall also that $\Gamma_n(p)$ is defined in (2.15). To avoid trivialities, in the following statements we assume that G is not complete in order to have at least one minimal separator. Theorems 3.3 and 3.4 are useful only for non homogeneous graphs, since stronger results, Theorems 3.1 and 3.2, are available for homogeneous graphs.

Theorem 3.3 *Let G be a non complete decomposable graph and let P be a perfect order of its cliques. Then $A_P \subset \mathcal{A}$. More specifically for $y \in P_G$ and for $(\alpha, \beta) \in A_P$ the integral (3.1) converges and*

$$\int_{Q_G} e^{-\langle x, y \rangle} H_G(\alpha, \beta; x) \mu_G(dx) = \Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \varphi(y)) \quad (3.19)$$

where

$$\Gamma_I(\alpha, \beta) = \Gamma_{s_2}(\alpha_1 + \delta_2) \frac{\Gamma_{c_1}(\alpha_1)}{\Gamma_{s_2}(\alpha_1)} \prod_{q=2}^k \frac{\Gamma_{c_q}(\alpha_q)}{\Gamma_{s_q}(\alpha_q)}. \quad (3.20)$$

Equivalently, if we write $y = \hat{\sigma}^{-1}$ with $\sigma \in Q_G$, (3.19) can be rewritten as

$$\int_{Q_G} e^{-\langle x, \hat{\sigma}^{-1} \rangle} H_G(\alpha, \beta; x) \mu_G(dx) = \Gamma_1(\alpha, \beta) H_G(\alpha, \beta; \sigma) \quad (3.21)$$

To study \mathcal{B} for a nonhomogeneous graph and give the normalizing constant of the type II Wishart, we now need to define, for a given P , the set B_P to be the set of (α, β) such that

1. $\sum_{j \in J(P, S)} (\alpha_j + \frac{1}{2}(c_j - s_j)) - \nu(S)\beta(S) = 0$, for all S different from S_2 ;
2. $-\alpha_q - \frac{1}{2}(c_q - s_q - 1) > 0$ for all $q = 2, \dots, k$ and $-\alpha_1 - \frac{1}{2}(c_1 - s_2 - 1) > 0$
3. $-\alpha_1 - \frac{1}{2}(c_1 - s_2 + 1) - \gamma_2 > \frac{s_2 - 1}{2}$ where $\gamma_2 = \sum_{j \in J(P, S_2)} (\alpha_j - \beta_2 + \frac{c_j - s_2}{2})$.

Theorem 3.4 *Let G be a noncomplete decomposable graph and let P be a perfect order of its cliques. Then $B_P \subset \mathcal{B}$. More specifically for $\theta \in Q_G$ and $(\alpha, \beta) \in B_P$ the integral (3.2) converges and*

$$\int_{P_G} e^{-\langle \theta, y \rangle} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) = \Gamma_{II}(\alpha, \beta) H_G(\alpha, \beta; \theta) \quad (3.22)$$

where

$$\Gamma_{II}(\alpha, \beta) = \Gamma_{s_2} \left[-\alpha_1 - \frac{c_1 - s_2}{2} - \gamma_2 \right] \frac{\Gamma_{c_1}(-\alpha_1)}{\Gamma_{s_2}(-\alpha_1 - \frac{c_1 - s_2}{2})} \prod_{j=2}^k \frac{\Gamma_{c_j}(-\alpha_j)}{\Gamma_{s_j}(-\alpha_j - \frac{c_j - s_j}{2})} \quad (3.23)$$

It is interesting to reexpress (3.22) in a slightly different way. Writing $y = \hat{x}^{-1}$ with $x \in Q_G$ and recalling that the image of $\nu_G(dy)$ under $y \mapsto \varphi(y) = x$ is $\mu_G(dx)$, we see that (3.22) can be rewritten

$$\int_{Q_G} e^{-\langle \theta, \hat{x}^{-1} \rangle} H_G(\alpha, \beta; x) \mu_G(dx) = \Gamma_{II}(\alpha, \beta) H_G(\alpha, \beta; \theta) \quad (3.24)$$

From the two theorems above, it follows immediately that

$$\mathcal{A} \supset \cup_P A_P \quad \text{and} \quad \mathcal{B} \supset \cup_P B_P$$

where the union of all A_P and all B_P is taken over all possible perfect order of the cliques of G . Before making some important remarks, let us give an example.

Example: Consider the graph $G = A_4 : \bullet^1 - \bullet^2 - \bullet^3 - \bullet^4$.

Let $C_1 = \{1, 2\}$, $C_2 = \{2, 3\}$, $C_3 = \{3, 4\}$, $S_2 = \{2\}$, $S_3 = \{3\}$, and let $\alpha(C_i) = \alpha_i$, $i = 1, 2, 3$, $\beta(S_i) = \beta_i$, $i = 2, 3$. Then $P_1 = (C_1, C_2, C_3)$ and $P_2 = (C_2, C_1, C_3)$ are perfect orders of the cliques. The orders $P'_1 = (C_3, C_2, C_1)$ and $P'_2 = (C_2, C_3, C_1)$ are also perfect orders analog respectively to P_1 and P_2 . On the other hand the only other possible orders (C_1, C_3, C_2) and its analog (C_3, C_1, C_2) are not perfect. Let us therefore identify A_P and B_P for P_1, P'_1 and P_2, P'_2 :

$$A_{P_1} = \{(\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3) \mid \alpha_i > \frac{1}{2}, i = 1, 2, 3, \alpha_1 + \alpha_2 - \beta_2 > 0, \alpha_3 = \beta_3\}$$

$$A_{P'_1} = \{(\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3) \mid \alpha_i > \frac{1}{2}, i = 1, 2, 3, \alpha_2 + \alpha_3 - \beta_3 > 0, \alpha_1 = \beta_2\}$$

while $A_{P_2} = A_{P_1}$ and $A_{P'_2} = A_{P'_1}$. In a parallel way, we have

$$B_{P_1} = \{(\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3) \mid -\alpha_i > 0, i = 1, 2, 3, -\alpha_1 - \alpha_2 + \beta_2 - 1 > 0, \alpha_3 + \frac{1}{2} = \beta_3\}$$

$$B_{P'_1} = \{(\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3) \mid -\alpha_i > 0, i = 1, 2, 3, -\alpha_2 - \alpha_3 + \beta_3 - 1 > 0, \alpha_1 + \frac{1}{2} = \beta_2\}$$

while $B_{P_2} = B_{P_1}$ and $B_{P'_2} = B_{P'_1}$.

Remarks.

1. The domains A_P and B_P on which (3.19) and (3.22) respectively or equivalently (3.21) and (3.24) hold, depend upon the chosen perfect order P of the cliques. Since the functions H_G do not depend upon P , it is clear that, even though the expression of Γ_I and Γ_{II} depend upon P , their values do not.
2. Since Assumption 1 of Theorems 3.3 and 3.4 represents $k' - 1$ constraints on the set of (α, β) 's, we see that in general each set A_P is of dimension $k + 1$.
3. From Theorem 3.3 and 3.4, the integrals (3.1) and (3.2) are finite and constant multiples of $H_G(\alpha, \beta; \sigma)$ and $H_G(\alpha, \beta; \theta)$ for (α, β) in $\cup_P A_P$ and $\cup_P B_P$ respectively. Using Hölder's inequality it is immediate to prove that these integrals are also finite on the convex hull of $\cup_P A_P$ and $\cup_P B_P$. So the question naturally arises as to whether \mathcal{A} and \mathcal{B} are larger than $\cup_P A_P$ and $\cup_P B_P$. We only have a partial answer to this. We have seen in the previous section that, when G is homogeneous, \mathcal{A} and \mathcal{B} are completely known, and of full dimension $k + k'$. However, if we consider the homogeneous example given in §3.3 and treat it using the methods given in this section, we will find that the 24 possible orders are all perfect with $P_1 = (C_1, C_2, C_3, C_4)$ and $P_2 = (C_1, C_4, C_2, C_3)$ being the only perfect orders yielding distinct A_P 's. We have

$$\begin{aligned}
A_{P_1} &= \{(\alpha, \beta) \mid \alpha_i > 1, i = 1, 2, 3, \alpha_4 > \frac{1}{2}, \alpha_1 + \alpha_2 + \alpha_3 - 2\beta_2 > \frac{1}{2}, \alpha_4 = \beta_1\}, \\
B_{P_1} &= \{(\alpha, \beta) \mid \alpha_i < 0, i = 1, 2, 3, 4, -\alpha_1 - \alpha_2 - \alpha_3 + 2\beta_2 > \frac{5}{2}, \alpha_4 - \beta_1 + \frac{1}{2} = 0\}, \\
A_{P_2} &= \{(\alpha, \beta) \mid \alpha_i > 1, i = 1, 2, 3, \alpha_4 > \frac{1}{2}, \alpha_1 + \alpha_4 - \beta_1 > 0, \alpha_2 + \alpha_3 - 2\beta_2 = 0\}, \\
B_{P_2} &= \{(\alpha, \beta) \mid \alpha_i < 0, i = 1, 2, 3, 4, -\alpha_1 - \alpha_4 + \beta_1 > 2, \alpha_2 + \alpha_3 - 2\beta_2 + 1 = 0\}.
\end{aligned}$$

Clearly, $A_{P_1} \cup A_{P_2}$ is included in, but not equal to, \mathcal{A} as given in (3.17). Similarly, $B_{P_1} \cup B_{P_2}$ is included in, but not equal to, \mathcal{B} as given in (3.18). The question is therefore whether in the nonhomogeneous case it is possible to identify \mathcal{A} and \mathcal{B} . In the next section, we find \mathcal{A} and \mathcal{B} for $G = A_4$ and we see that they are of dimension strictly less than $k + k'$. Thus $G = A_4$ is a counterexample to the hypothesis that in the nonhomogeneous case, we could also define a set of dimension $k + k'$ on which (3.21) and (3.22) hold.

3.5 The case $G = A_4$

Let G be A_4 as in the previous example. Then we write

$$\sigma = \begin{pmatrix} \sigma_1 & \sigma_{12} & & & \\ \sigma_{21} & \sigma_2 & \sigma_{23} & & \\ & \sigma_{32} & \sigma_3 & \sigma_{34} & \\ & & \sigma_{43} & \sigma_4 & \end{pmatrix}$$

for $\sigma \in Q_G$, with $\sigma_{ij} = \sigma_{ji}$, $\sigma_{i,j} = \sigma_i - \sigma_{ij}\sigma_j^{-1}\sigma_{ji}$ and similarly for $\theta \in Q_G$.

Proposition 3.2 *Consider the graph $G = A_4$ with cliques and separators*

$$C_1 = \{1, 2\}, C_2 = \{2, 3\}, C_3 = \{3, 4\}, S_2 = \{2\}, S_3 = \{3\}.$$

Let $\alpha_i = \alpha(C_i), i = 1, 2, 3$ $\beta_i = \beta(S_i), i = 2, 3$. Define

$$\mathcal{A}_4 = \{(\alpha, \beta) \mid \alpha_i > \frac{1}{2}, i = 1, 2, 3, \alpha_1 + \alpha_2 > \beta_2, \alpha_2 + \alpha_3 > \beta_3\}.$$

Then the following integral converges for all $\sigma \in Q_{A_4}$ if and only if (α, β) is in \mathcal{A}_4 . Under these conditions, it is equal to

$$\begin{aligned}
& \int_{Q_G} e^{-(x, \hat{\sigma}^{-1})} H_G(\alpha, \beta; x) \mu_G(dx) \tag{3.25} \\
&= \pi^{\frac{3}{2}} \frac{\Gamma(\alpha_1 - \frac{1}{2}) \Gamma(\alpha_2 - \frac{1}{2}) \Gamma(\alpha_3 - \frac{1}{2}) \Gamma(\alpha_1 + \alpha_2 - \beta_2) \Gamma(\alpha_2 + \alpha_3 - \beta_3)}{\Gamma(\alpha_2)} \\
& \quad \sigma_{1,2}^{\alpha_1} \sigma_{2,3}^{\alpha_1 + \alpha_2 - \beta_2} \sigma_{3,2}^{\alpha_2 + \alpha_3 - \beta_3} \sigma_{4,3}^{\alpha_3} {}_2F_1(\alpha_1 + \alpha_2 - \beta_2, \alpha_2 + \alpha_3 - \beta_3, \alpha_2, \frac{\sigma_{23}^2}{\sigma_2 \sigma_3})
\end{aligned}$$

where ${}_2F_1$ denotes the hypergeometric function. Similarly we define

$$\mathcal{B}_4 = \{(\alpha, \beta) \mid \alpha_1 < 0, \alpha_3 < 0, \beta_2 - \alpha_1 - \alpha_2 - \frac{1}{2} > 0, \beta_3 - \alpha_2 - \alpha_3 - \frac{1}{2} > 0, \beta_2 + \beta_3 - \alpha_1 - \alpha_2 - \alpha_3 - \frac{3}{2} > 0\}.$$

Then the following integral converges for all for $\theta \in Q_G$ if and only if $(\alpha, \beta) \in \mathcal{B}_4$. Under these conditions, it is equal to

$$\begin{aligned} & \int_{P_G} e^{-(\theta; y)} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) \\ &= \pi^{\frac{3}{2}} \theta_{1.2}^{\alpha_1} \theta_2^{\alpha_1 + \alpha_2 - \beta_2} \theta_3^{\alpha_2 + \alpha_3 - \beta_3} \theta_{4.3}^{\alpha_3} \\ & \quad \frac{\Gamma(-\alpha_1) \Gamma(\beta_2 - \alpha_1 - \alpha_2 - \frac{1}{2}) \Gamma(\beta_2 + \beta_3 - \alpha_1 - \alpha_2 - \alpha_3 - \frac{3}{2}) \Gamma(\beta_3 - \alpha_2 - \alpha_3 - \frac{1}{2}) \Gamma(-\alpha_3)}{\Gamma(\beta_2 + \beta_3 - \alpha_1 - \alpha_2 - \alpha_3 - 1)} \\ & \quad {}_2F_1\left(\beta_2 - \alpha_1 - \alpha_2 - \frac{1}{2}, \beta_3 - \alpha_2 - \alpha_3 - \frac{1}{2}; \beta_2 + \beta_3 - \alpha_1 - \alpha_2 - \alpha_3 - 1; \frac{\theta_{23}^2}{\theta_2 \theta_3}\right). \end{aligned} \tag{3.26}$$

The results above are obtained by a non trivial and long computation. A central part of this computation is the following lemma.

Lemma 3.1 Consider the following 2×2 random matrix $X = \begin{bmatrix} X_1 & X_{12} \\ X_{12} & X_2 \end{bmatrix}$ with the Wishart distribution

$$w_2(p, c^{-1}; dx) = \frac{(\det c)^p}{\Gamma_2(p)} e^{-\langle x, c \rangle} (x_1 x_2 - x_{12}^2)^{p - \frac{3}{2}} \mathbf{1}_{M_2^+}(x) dx_1 dx_2 dx_{12}$$

with $p \geq 1/2$ and $c = \begin{bmatrix} c_1 & c_{12} \\ c_{12} & c_2 \end{bmatrix}$ positive definite. For $a_1 > -p$ and $a_2 > -p$, the Mellin transform of (X_1, X_2) is

$$E(X_1^{a_1} X_2^{a_2}) = \frac{(\det c)^p}{c_1^{a_1+p} c_2^{a_2+p}} \frac{\Gamma(a_1 + p) \Gamma(a_2 + p)}{\Gamma(p)^2} {}_2F_1(a_1 + p, a_2 + p; p; \frac{c_{12}^2}{c_1 c_2})$$

The proofs of Proposition 3.2 and Lemma 3.1 are omitted.

We now derive from Proposition 3.2 the sets \mathcal{A} and \mathcal{B} when $G = A_4$.

Corollary 3.1 Let $G = A_4$. Then $\mathcal{A} = \cup A_P$ and $\mathcal{B} = \cup B_P$, where the unions are taken over the two possible A_P and B_P . The dimension of \mathcal{A} and \mathcal{B} is therefore strictly less than $k + k'$.

Proof: Since the two statements are quite similar, we prove the second one only. We use the equality (see Abramowitz and Stegun, formula 15.3.3)

$$(1 - z)^{a+b-c} {}_2F_1(a, b; c; z) = {}_2F_1(c - a, c - b; c; z). \tag{3.27}$$

Using (3.27) in the right hand side of (3.26) above with $z = \frac{\theta_{23}^2}{\theta_2\theta_3}$ and

$$(a, b, c) = (\beta_2 - \alpha_1 - \alpha_2 - \frac{1}{2}, \beta_3 - \alpha_2 - \alpha_3 - \frac{1}{2}; \beta_2 + \beta_3 - \alpha_1 - \alpha_2 - \alpha_3 - 1)$$

we see that

$$\frac{1}{H_G(\alpha, \beta; \theta)} \int_{Q_G} e^{-\langle \theta, y \rangle} H_G(\alpha, \beta; \varphi(y)) \nu_G(dy) = C {}_2F_1(c - a, c - b; c; z)$$

where C is a constant which does not depend on z . Now clearly from its Taylor expansion the hypergeometric function $z \mapsto {}_2F_1(c - a, c - b; c; z)$ is a constant if and only if either $c - a$ or $c - b$ is zero, which together with \mathcal{B}_4 proves that constancy occurs if and only if (α, β) belongs to one of the two possible B_P as given in the example of §3.4 above. \square

Remark. One can prove that the convex hulls in \mathbb{R}^5 of \mathcal{A} and \mathcal{B} are respectively strictly included in \mathcal{A}_4 and \mathcal{B}_4 as defined in Proposition 3.2.

4 Properties of the type I and II Wisharts

Let us recall that for a given decomposable graph G , the r -dimensional graphical Gaussian model Markov with respect to G is the family of distributions

$$\mathcal{N}_G = \{N_r(0, \Sigma), \Sigma \in Q_G\}.$$

Dawid and Lauritzen (1993, p. 1306) have shown that this model is strong meta Markov. This can also be shown directly since, using the notation of (2.8), for a given perfect order of the cliques and with the convention that $x_{[1]} = x_{C_1}$, $\Sigma_{[1]} = \Sigma_{C_1}$, $x_{\langle 1 \rangle} = 0$, $r_1 = c_1$, the density of $X \in \mathbb{R}^r$ with distribution $N_r(0, \Sigma) \in \mathcal{N}_G$ can be written as

$$f(x) = \prod_{i=1}^k \frac{1}{(2\pi)^{\frac{r_i}{2}} \Sigma_{[i]}^{\frac{1}{2}}} e^{-\frac{1}{2} \langle (x_{[i]} - x_{[i]_{<i>}} \Sigma_{<i>}^{-1} \Sigma_{<i>}, \Sigma_{[i]}^{-1} (x_{[i]} - x_{[i]_{<i>}} \Sigma_{<i>}^{-1} \Sigma_{<i>}) \rangle} \quad (4.1)$$

The parameters

$$F_1 = \Sigma_{C_1}, (L_i = \Sigma_{[i]_{>}} \Sigma_{<i>}^{-1}, N_i = \Sigma_{[i]}), i = 2, \dots, k \quad (4.2)$$

of the distributions of $X_{C_1}, X_{[i]} | x_{H_{i-1}}, i = 2, \dots, k$ respectively are clearly variation independent in the sense that any parameter $(\Sigma_{[i]_{>}} \Sigma_{<i>}^{-1}, \Sigma_{[i]})$ of the distribution of $X_{[i]} | X_{H_{i-1}}$ is compatible with any parameter $\{F_1, (L_j, N_j), j < i\}$ of $X_{H_{i-1}}$ and any parameter $\{L_j, N_j, j > i\}$ of $X | X_{H_i}$. It then follows that for a decomposition (A, B) of G the parameter Σ_A of the distribution of X_A is variation independent of the parameter $\Sigma_{B|A}$ of the conditional distribution of X_B given X_A . This, according to Definition 4.3 of Dawid and Lauritzen (1993),

means that the model \mathcal{N}_G is strong meta-Markov. For this model, Dawid and Lauritzen (1993, p. 1306 and p. 1308) have shown that the distribution of the maximum likelihood estimator of Σ , that is the hyper Wishart, is weak hyper Markov and that the hyper inverse Wishart, the inverse of the G -Wishart, is a conjugate prior on Σ which is strong hyper Markov.

We are now going to show parallel results for the type I and II Wishart: the type I Wishart is weak hyper Markov, the inverse of the type II Wishart forms a conjugate family for the scale parameter of the \mathcal{N}_G model and for any direction given to the graph by a perfect order of its cliques, the inverse type II Wishart is strong directed hyper Markov. Since we have seen in §3.2 that the hyper Wishart is a particular case of the type I Wishart and the hyper inverse Wishart is a particular case of the inverse type II, it is not surprising that their generalizations hold parallel Markov properties.

One might wonder whether the term “hyper” is adequate when talking about the weak Markov property of the type I Wishart since this distribution has been identified so far neither as the distribution of an estimator nor as a prior distribution for the parameter of a Gaussian model. It is certainly adequate for the inverse of the type II Wishart since, as we are going to prove right away in §4.1, it forms a conjugate family of prior distributions for the scale parameter of the \mathcal{N}_G model, with a shape parameter set of dimension at least $k + 1$. We will then prove the hyper Markov properties in §4.2. Our main results below are Corollary 4.1 and Theorem 4.4.

4.1 Conjugate prior distributions

The family of inverse type II Wishart distributions has several properties that make it useful as a rich family of conjugate prior distributions for the scale parameter Σ_G of the graphical Gaussian model Markov with respect to a decomposable graph G . Recall that, following the notation used in the introduction, if Σ is the positive definite covariance matrix for $N_r(0, \Sigma) \in \mathcal{N}_G$, then $\Sigma_G = \pi(\Sigma) \in Q_G$ is the scale parameter for the $N_r(0, \Sigma)$ distribution. We have the following general result.

Theorem 4.1 *Let G be a decomposable graph and let P be a perfect order of its cliques. Let D be in Q_G , let (α, β) be in A_P and (α', β') be in B_P . If the joint distribution of (X, Σ_G) on $Q_G \times Q_G$ is $W_{Q_G}(\alpha, \beta, \sigma; dx)IW_{P_G}(\alpha', \beta', D; d\sigma)$, then the conditional distribution of Σ_G knowing $X = x$ is $IW_{P_G}(\alpha' - \alpha, \beta' - \beta, D + x; d\sigma)$ and the marginal distribution of X is an F distribution of the first kind with parameter $(\alpha, \beta, \alpha', \beta', D)$.*

Proof: The joint distribution of (X, Σ_G) is

$$W_{Q_G}(\alpha, \beta, \sigma; dx)IW_{P_G}(\alpha', \beta', D; d\sigma)$$

$$\begin{aligned}
&= \frac{e^{-\langle x, \hat{\sigma}^{-1} \rangle} H_G(\alpha, \beta; x)}{\Gamma_I(\alpha, \beta) H_G(\alpha, \beta; \sigma)} \mu_G(dx) \frac{e^{-\langle D, \hat{\sigma}^{-1} \rangle} H_G(\alpha', \beta'; \sigma)}{\Gamma_{II}(\alpha', \beta') H_G(\alpha', \beta'; D)} \mu_G(d\sigma) \\
&= \left[\frac{H_G(\alpha, \beta; x)}{\Gamma_I(\alpha, \beta) \Gamma_{II}(\alpha', \beta') H_G(\alpha', \beta'; D)} \mu_G(dx) \right] e^{-\langle x+D, \hat{\sigma}^{-1} \rangle} H_G(\alpha' - \alpha, \beta' - \beta; \sigma) \mu_G(d\sigma)
\end{aligned}$$

from which the result follows immediately. \square

Theorem 4.1 shows that the family of IW_{P_G} distributions is a conjugate family for the scale parameter σ of the $W_{Q_G}(\alpha, \beta, \sigma; dx)$. Consider now a sample Z_1, \dots, Z_n from a Gaussian distribution Markov with respect to G , let $S = \frac{1}{n} \sum_{i=1}^n Z_i Z_i^t$. Then $\pi(S)$, the maximum likelihood estimator of Σ_G is such that $n\pi(S)$ is hyper Wishart with shape parameter $p = \frac{n}{2}$ and scale parameter Σ_G , that is, Wishart of Type I with shape parameter

$$\alpha(C) = p, \quad C \in \mathcal{C}, \quad \beta(S) = p, \quad S \in \mathcal{S}, \quad p > \max_{C \in \mathcal{C}} \frac{|C| - 1}{2}$$

and scale parameter $2\Sigma_G$. Applying Theorem 4.1 to $X = n\pi(S)$, we obtain the following corollary

Corollary 4.1 *Let G be decomposable and let P be a perfect order of its cliques. Let (Z_1, \dots, Z_n) be a sample from the $N_r(0, \Sigma)$ distribution with $\Sigma_G \in Q_G$. If the prior distribution on $2\Sigma_G$ is $IW_{P_G}(\alpha', \beta', D)$ with $(\alpha', \beta') \in B_P$ and $D \in Q_G$, the posterior distribution of $2\Sigma_G$, given $nS = \sum_{i=1}^n Z_i Z_i^t$ is $IW_{P_G}(\alpha' - \frac{n}{2}, \beta' - \frac{n}{2}, D + \pi(nS))$, where $\alpha' - \frac{n}{2} = (\alpha_1 - \frac{n}{2}, \dots, \alpha_k - \frac{n}{2})$ and $\beta' - \frac{n}{2} = (\beta'_1 - \frac{n}{2}, \dots, \beta'_k - \frac{n}{2})$ are such that $(\alpha' - \frac{n}{2}, \beta' - \frac{n}{2}) \in B_P$ and $D + \pi(nS) \in Q_G$.*

This means that the family $\{IW_{P_G}(\alpha, \beta, D), (\alpha, \beta) \in \mathcal{A}, D \in Q_G\}$ is a conjugate family for the scale parameter Σ_G of the Gaussian model Markov with respect to G . We note this family has its shape parameter set of dimension at least $k+1$ and is therefore much richer than the traditional Diaconis-Ylvisaker family with shape parameter set of dimension equal to 1.

Theorem 4.1 can also be immediately transcribed to the homogeneous case using the variables $(x_{[u]}, x_{[u>}x_{<u}^{-1}), u \in T)$ and we obtain the following result.

Theorem 4.2 *Let G be a homogeneous graph. Let D be in Q_G , let (α, β) be in \mathcal{A} and (α', β') be in \mathcal{B} . If $(X, \Sigma_G) \in Q_G \times Q_G$ and the joint distribution of $(X_{[u]}, X_{[u>}X_{<u}^{-1}, \Sigma_{[u]}, \Sigma_{[u>} \Sigma_{<u}^{-1}), u \in T)$ is $W_{Q_G}^*(\alpha, \beta, \sigma) IW_{P_G}^*(\alpha', \beta', D)$, then the conditional distribution of Σ knowing $X = x$ is $IW_{P_G}^*(\alpha' - \alpha, \beta' - \beta, D + x)$.*

We now have the following result dual to Theorem 4.1.

Theorem 4.3 *Let P be a perfect order of the cliques of G . Let σ be in Q_G . Let (α, β) be in B_P and (α', β') be in A_P . If the joint distribution of (Y, Θ) on $P_G \times P_G$ is $W_{P_G}(\alpha, \beta, \theta)IW_{Q_G}(\alpha', \beta', \sigma)$, then the conditional distribution of Θ knowing $Y = y$ is $IW_{Q_G}(\alpha' - \alpha, \beta' - \beta, \varphi(y + \hat{\sigma}^{-1}))$.*

Theorem 4.3 shows that the family of IW_{Q_G} distributions is a conjugate family for the scale parameter θ of the $W_{P_G}(\alpha, \beta, \theta)$.

4.2 Markov properties

We now want to show that the type II inverse Wishart IW_{P_G} is strong directed hyper Markov and the type I Wishart weak hyper Markov. Let $M(G)$ denote the set of all Markov probabilities over G . Let the distribution $P_\rho \in M(G)$ be parametrized by ρ . Now we randomize ρ according to a law $\mathcal{L}(\rho)$. For any subset A of V , let ρ_A denote the parameter of the marginal distribution of X_A (more specifically we should write $\rho \sim_A \rho'$ if the marginal distributions of X_A under P_ρ and $P_{\rho'}$ coincide and call ρ_A the equivalence class of ρ for the equivalence relation \sim_A .) The parameter $\rho_{A|B}$ of the conditional distribution of X_A knowing X_B could be defined in a similar way. We say that $\mathcal{L}(\rho)$ is weak hyper Markov over G if under $\mathcal{L}(\rho)$, for any decomposition (A, B) of V ,

$$\rho_A \perp\!\!\!\perp \rho_B | \rho_{A \cap B}. \quad (4.3)$$

We say that $\mathcal{L}(\rho)$ is strong hyper Markov over G if, under $\mathcal{L}(\rho)$, for any decomposition (A, B) of V ,

$$\rho_{A|B} \perp\!\!\!\perp \rho_B. \quad (4.4)$$

Let P be any perfect order of the cliques and consider a perfect numbering of the vertices compatible with P (see Lauritzen (1996), p. 18). Let D be the directed graph obtained from G by directing all edges in G from the vertex with the smallest number to the vertex with the highest number. We say that a law $\mathcal{L}(\rho)$ is weak directed hyper Markov over D if for all $v \in V$,

$$\rho_v \perp\!\!\!\perp \rho_{pr(v)} | \rho_{pa(v)}, \quad (4.5)$$

where $pa(v)$ denotes the sets of parents of v in D and $pr(v)$ denotes the sets of predecessors, i.e. the vertices with a lower number than v .

We say that $\mathcal{L}(\rho)$ is strong directed hyper Markov over D if for all $v \in V$,

$$\rho_{v|pa(v)} \perp\!\!\!\perp \rho_{pr(v)}. \quad (4.6)$$

Let us also recall that a random variable on M_r^+ is said to follow the Wishart $w_r(p, \sigma)$ distribution if its density with respect to the Lebesgue measure is

$$\frac{|x|^{p-\frac{r+1}{2}}}{|\sigma|^p \Gamma_r(p)} e^{-\langle x, \sigma^{-1} \rangle},$$

in which case its inverse $U = X^{-1}$ is said to follow the inverse Wishart distribution $iw_r(p, \theta)$, where $\theta = \sigma^{-1}$, with density with respect to the Lebesgue measure, equal to

$$\frac{|u|^{-p-\frac{r+1}{2}}}{|\sigma|^p \Gamma_r(p)} e^{-\langle u^{-1}, \theta \rangle}.$$

Finally, we will use the notation $x_{[12>}$ and $x_{[1]}$ for

$$x_{[12>} = x_{C_1 \setminus S_2, S_2} x_{S_2}^{-1} \quad \text{and} \quad x_{[1]} = x_{C_1 \setminus S_2, S_2}.$$

Theorem 4.4 *Let G be a decomposable graph G and let P be a perfect order of its cliques. Then, for $(\alpha, \beta) \in B_P$ and for the direction given by P , the inverse type II Wishart is strong directed hyper Markov. More precisely, if $X \sim IW_{P_G}(\alpha, \beta, \theta)$ with $(\alpha, \beta) \in B_P$ and $\theta \in Q_G$, then*

$$\begin{aligned} x_{[i]} &\sim iw_{c_i - s_i}(-\alpha_i, \theta_{[i]}), \quad i = 1, \dots, k \\ x_{[12>} | x_{[1]} &\sim N_{(c_1 - s_2) \times s_2}(\theta_{[12>}, 2 \theta_{<2>}^{-1} \otimes x_{[1]}) \\ x_{<2>} &\sim iw_{s_2}(-(\alpha_1 + \frac{c_1 - s_2}{2} + \gamma_2), \theta_{<2>}) \\ x_{[j>} x_{<j>}^{-1} | x_{[j]} &\sim N_{(c_j - s_j) \times s_j}(\theta_{[j>} \theta_{<j>}^{-1}, 2 \theta_{<j>}^{-1} \otimes x_{[j]}), \quad j = 2, \dots, k \end{aligned}$$

and

$$\{(x_{[12>}, x_{[1]}), x_{<2>}, (x_{[j>} x_{<j>}^{-1}, x_{[j]}), j = 2, \dots, k\} \quad (4.7)$$

are mutually independent.

Proof: From (4.1), we know that $(\Sigma_{[i>} \Sigma_{<i>}^{-1}, \Sigma_{[i]})$ is the parameter of the distribution of $Z_{[i]}$ given $Z_{<i>}$ when $Z \sim N_r(0, \Sigma) \in \mathcal{N}_G$. Therefore it follows from the remark following Theorem 2.6 and Proposition 3.8 of Dawid and Lauritzen (1993) that to construct a weak hyper Markov distribution for $\rho = \Sigma_G \in Q_G$, it is sufficient to build a weak directed hyper Markov distribution for a given direction of the vertices compatible with a given perfect order of the cliques, that is a distribution with density of the form

$$\begin{aligned} p(\Sigma_G) &= p_{C_1}(\Sigma_{C_1}) \prod_{i=2}^k p_i(\Sigma_{[i>} \Sigma_{<i>}^{-1}, \Sigma_{[i]} | \Sigma_{<j>}) \\ &= p_1(\Sigma_{[12>}, \Sigma_{[1]} | \Sigma_{<2>}) p_{<2>}(\Sigma_{<2>}) \prod_{i=2}^k p_i(\Sigma_{[i>} \Sigma_{<i>}^{-1}, \Sigma_{[i]} | \Sigma_{<j>}). \quad (4.8) \end{aligned}$$

If we want to show that, for the given direction, this distribution is in fact strong directed hyper Markov, by Proposition 3.13 of Dawid and Lauritzen (1993) it is sufficient to show that

$$\{(\Sigma_{[12>}, \Sigma_{[1\cdot]}, \Sigma_{<2>}, (\Sigma_{[j>} \Sigma_{<j>}^{-1}, \Sigma_{[j\cdot]}, j = 2, \dots, k)\} \quad (4.9)$$

are mutually independent. Let us now show that the Inverse type II Wishart satisfies both (4.8) and (4.9). Let $X \sim IW_{P_G}(\alpha, \beta, \theta)$ with $(\alpha, \beta) \in B_P$ and $\theta \in Q_G$. Combining (7.9) and (7.11) of the appendix, we show that the image of the $IW_{P_G}(\alpha, \beta, \theta; dx)$ distribution by the change of variables (2.12) and (7.5) is

$$IW_{P_G}^{**}(\alpha, \beta, \theta; d(x_{[1\cdot]}, d(x_{[12>}, d(x_{S_2}), d(x_{[j>} x_{<j>}^{-1}), dx_{[j\cdot]}, j = 2, \dots, k) \\ \propto e^{-\langle x_{[1\cdot]}^{-1}, \theta_{[1\cdot]} \rangle} |x_{[1\cdot]}|^{\alpha_1 - \frac{c_1 - s_2 + 1}{2}} |x_{[1\cdot]}|^{-\frac{s_2}{2}} e^{-\langle (x_{[12>} - \theta_{[12>}), x_{[1\cdot]}^{-1} (x_{[12>} - \theta_{[12>}) \theta_{<2>} \rangle} \quad (4.10)$$

$$e^{-\langle x_{<2>}^{-1}, \theta_{<2>} \rangle} |x_{<2>}|^{\alpha_1 + \frac{c_1 - s_2}{2} + \gamma_2 - \frac{s_2 + 1}{2}} dx_{<2>} \quad (4.11)$$

$$\prod_{j=2}^k |x_{[j\cdot]}|^{\alpha_j - \frac{c_j - s_j + 1}{2}} e^{-\langle x_{[j\cdot]}^{-1}, \theta_{[j\cdot]} \rangle} |x_{[j\cdot]}|^{-\frac{s_j}{2}} \quad (4.12)$$

$$\prod_{j=2}^k e^{-\langle (x_{[j>} x_{<j>}^{-1} - \theta_{[j>} \theta_{<j>}^{-1}), x_{[j\cdot]}^{-1} (x_{[j>} x_{<j>}^{-1} - \theta_{[j>} \theta_{<j>}^{-1}) \theta_{<j>} \rangle} \quad (4.13)$$

$$dx_{[12>} dx_{[1\cdot]} \prod_{j=2}^k d(x_{[j>} x_{<j>}^{-1}) dx_{[j\cdot]}.$$

We see that the densities (4.10), (4.11), (4.12) and (4.13) with x replaced by Σ_G are exactly of the form required for the respective factors of (4.8). It follows that the $IW_{P_G}(\alpha, \beta, \theta)$ is weak directed hyper Markov but we also see from (4.10) - (4.13) above that the independence in (4.9) is satisfied and therefore $IW_{P_G}(\alpha, \beta, \theta)$ is strong directed hyper Markov. The densities of

$$(x_{[1\cdot]}, x_{[12>}, x_{<2>}, x_{[j\cdot]}, x_{[j>} x_{<j>}^{-1}, j = 2, \dots, k)$$

are also clearly as indicated in the theorem. \square

This strong directed hyper Markov property of the type II inverse Wishart corresponds to the strong hyper Markov property for the inverse G -Wishart, i.e. the hyper inverse Wishart. We do not quite have the strong hyper-Markov property because the parameters $(\alpha, \beta) \in B_P$ are linked to the perfect order P . The property analog to the weak hyper Markov property of the hyper Wishart is given in the following theorem.

Theorem 4.5 *Let G be a decomposable graph and let P be a perfect order of its cliques. Then, for $(\alpha, \beta) \in A_P$, the type I Wishart is weak hyper Markov. More precisely, if $X \sim W_{Q_G}(\alpha, \beta, \sigma)$ with $(\alpha, \beta) \in A_P$ and $\sigma \in Q_G$, then*

$$x_{[1\cdot]} \sim w_{c_1 - s_2}(\alpha_1 - \frac{s_2}{2}, \sigma_{[1\cdot]})$$

$$\begin{aligned}
x_{[12>}|x_{<2>} &\sim N_{(c_1-s_2)\times s_2}(\sigma_{[12>}, 2 x_{<2>}^{-1} \otimes \sigma_{[1]}) \\
x_{<2>} &\sim w_{s_2}(\alpha_1 + \delta_2, \sigma_{<2>}) \\
x_{[j>}x_{<j>}^{-1}|x_{<j>} &\sim N_{(c_j-s_j)\times s_j}(\sigma_{[j>}\sigma_{<j>}^{-1}, 2 x_{<j>}^{-1} \otimes \sigma_{[j]}) \\
x_{[j]} &\sim w_{c_j-s_j}(\alpha_j - \frac{s_j}{2}, \sigma_{[j]}), j = 2, \dots, k
\end{aligned}$$

Proof: Using (7.4) and (7.6) of the Appendix, we see that the image of the $W_{Q_G}(\alpha, \beta, \sigma; dx)$ distribution by the change of variables (2.12) and (7.5) is

$$\begin{aligned}
&W_{Q_G}^{**}(\alpha, \beta, \sigma; dx_{[1]}, dx_{[12>}, dx_{<2>}, dx_{[j]}, d(x_{[j>}x_{<j>}^{-1}), j = 2, \dots, k) \\
&\propto |x_{[1]}|^{\alpha_1 - \frac{s_2}{2} - \frac{c_1 - s_2 + 1}{2}} e^{-\langle x_{[1]}, \sigma_{[1]}^{-1} \rangle} dx_{[1]}. \tag{4.14}
\end{aligned}$$

$$\times |x_{<2>}|^{\frac{c_1 - s_2}{2}} e^{-\langle (x_{[12>} - \sigma_{[12>})x_{<2>} > (x_{<2,1]} - \sigma_{<2,1]} \sigma_{[1]}^{-1} \rangle} dx_{[12>} \tag{4.15}$$

$$\times |x_{<2>}|^{\alpha_1 + \delta_2 - \frac{s_2 + 1}{2}} e^{-\langle x_{<2>}, \sigma_{<2>}^{-1} \rangle} dx_{<2>} \tag{4.16}$$

$$\times \prod_{j=2}^k |x_{[j]}|^{\alpha_j - \frac{s_j}{2} - \frac{c_j - s_j + 1}{2}} e^{-\langle x_{[j]}, \sigma_{[j]}^{-1} \rangle} dx_{[j]}. \tag{4.17}$$

$$\times |x_{<j>}|^{\frac{c_j - s_j}{2}} e^{-\langle (x_{[j>}x_{<j>}^{-1} - \sigma_{[j>}\sigma_{<j>}^{-1})x_{<j>} > (x_{<j>}^{-1}x_{<j]} - \sigma_{<j>}^{-1}\sigma_{<j]} \sigma_{[j]}^{-1} \rangle} d(x_{[j>}x_{<j>}^{-1}) \tag{4.18}$$

From the expression of the density $W_{Q_G}^{**}(\alpha, \beta, \sigma)$ above, we see that, for $x = \Sigma_G$, (4.14) and (4.15) give $p_1(\Sigma_{[C_1 \setminus S_2, S_2] > \Sigma_{S_2}^{-1}}, \Sigma_{C_1 \setminus S_2, S_2} | \Sigma_{S_2})$ of (4.8) while (4.16) gives $p_{S_2}(\Sigma_{S_2})$ and (4.17) and (4.18) give $p_i(\Sigma_{[i > \Sigma_{<i>}^{-1}}, \Sigma_{[i]} | \Sigma_{<j>})$. Therefore the $W_{Q_G}(\alpha, \beta, \sigma)$ type I Wishart is weak directed hyper Markov and, by Proposition 3.8 of Dawid and Lauritzen (1993), weak hyper Markov. \square

We note that in the proof above, the density $p_1(\Sigma_{[C_1 \setminus S_2, S_2] > \Sigma_{S_2}^{-1}}, \Sigma_{C_1 \setminus S_2, S_2} | \Sigma_{S_2})$ depends upon Σ_{S_2} and the density $p_i(\Sigma_{[i > \Sigma_{<i>}^{-1}}, \Sigma_{[i]} | \Sigma_{<j>})$ depends upon $\Sigma_{<j>}$ and therefore the $W_{Q_G}(\alpha, \beta, \sigma)$ type I Wishart is not strong directed hyper Markov.

4.3 Laplace transforms and expected values

For $(\alpha, \beta) \in \mathcal{A}$, $\mathcal{F}_{(\alpha, \beta), I} = \{W_{Q_G}(\alpha, \beta, \sigma; dx), \sigma \in Q_G\}$ is the natural exponential family generated by the measure

$$\mu_{(\alpha, \beta), G}(dx) = \frac{H_G(\alpha, \beta; x)}{\Gamma_I(\alpha, \beta)} \mu_G(dx).$$

For $-y \in P_G$, the Laplace transform of $\mu_{(\alpha, \beta), G}$ is

$$L_{\mu_{(\alpha, \beta), G}}(y) = \int_{Q_G} e^{\text{tr}(xy)} \mu_{(\alpha, \beta), G}(dx) = H_G(\alpha, \beta; -\varphi(y)).$$

This is a reformulation of (3.19). It implies that, for $-y + \hat{\sigma}^{-1} \in P_G$, the Laplace transform of $W_{Q_G}(\alpha, \beta, \sigma)$ is defined by

$$\int_{Q_G} e^{\text{tr}(xy)} W_{Q_G}(\alpha, \beta, \sigma; dx) = \frac{H_G(\alpha, \beta; \varphi(\hat{\sigma}^{-1} - y))}{H_G(\alpha, \beta; \sigma)}. \tag{4.19}$$

Suppose that (α, β) and (α', β') in \mathcal{A} are such that $(\alpha + \alpha', \beta + \beta')$ is still in \mathcal{A} . This is, for example, true for any G if (α, β) and (α', β') are in the same A_P and it is always true if G is homogeneous. We then have the convolution formula

$$W_{Q_G}(\alpha, \beta, \sigma) * W_{Q_G}(\alpha', \beta', \sigma) = W_{Q_G}(\alpha + \alpha', \beta + \beta', \sigma),$$

a result which would be difficult to prove using densities alone, *i.e.* without Theorem 3.3. Let us also mention some properties of the NEF $\mathcal{F}_{(\alpha, \beta), I}$. From Theorem 2.1 (1), we deduce that $L_{\mu_{(\alpha, \beta), G}}(y)$ is finite if and only if $-y \in P_G$. Since $-P_G$ is an open subset of Z_G (see Letac and Massam, 2006), $\mathcal{F}_{(\alpha, \beta), I}$ is a regular family in the sense of Barndorff Nielsen (1978) and the domain of the means $M_{\mathcal{F}_{(\alpha, \beta), I}}$ of the family $\mathcal{F}_{(\alpha, \beta), I}$ coincides with the interior of the closed convex support of the family. Thus $M_{\mathcal{F}_{(\alpha, \beta), I}} = Q_G$. The cumulant function of $\mu_{(\alpha, \beta), G}$ is

$$k_{\mu_{(\alpha, \beta), G}}(y) = \sum_{j=1}^k \alpha_j \log \det((-y^{-1})_{C_j}) - \sum_{j=2}^k \beta_j \log \det((-y^{-1})_{S_j}). \quad (4.20)$$

The computation of its differential requires some care and it is done in the Appendix in Propositions 7.1 and 7.2. We give the result here. If $X \sim W_{Q_G}(\alpha, \beta, \sigma)$, for $y = \hat{\sigma}^{-1} \in P_G$,

$$\begin{aligned} E_{W_{Q_G}(\alpha, \beta, \sigma)}(X) &= \frac{d}{dy} k_{\mu_{(\alpha, \beta), G}}(y) \\ &= \sum_{j=1}^k \alpha_j \tau(\hat{\sigma}_{V \setminus C_j, C_j} \sigma_{C_j}^{-1}) \sigma_{C_j} - \sum_{j=2}^k \beta_j \tau(\hat{\sigma}_{V \setminus S_j, S_j} \sigma_{S_j}^{-1}) \sigma_{S_j}, \end{aligned} \quad (4.21)$$

where

$$\begin{aligned} \tau(\hat{\sigma}_{V \setminus C_j, C_j} \sigma_{C_j}^{-1}) \sigma_{C_j} &= \begin{pmatrix} I_{C_j} & 0 \\ \hat{\sigma}_{V \setminus C_j, C_j} \sigma_{C_j}^{-1} & I_{V \setminus C_j} \end{pmatrix} \begin{pmatrix} \sigma_{C_j} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I_{C_j} & \sigma_{C_j}^{-1} \tau \sigma_{C_j, V \setminus C_j} \\ 0 & I_{V \setminus C_j} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{C_j} & \hat{\sigma}_{C_j, V \setminus C_j} \\ \hat{\sigma}_{V \setminus C_j, C_j} & \hat{\sigma}_{V \setminus C_j} \sigma_{C_j}^{-1} \sigma_{C_j, V \setminus C_j} \end{pmatrix} \\ &= \hat{\sigma} - \hat{\sigma}_{V \setminus C_j, C_j} \end{aligned} \quad (4.22)$$

with a similar expression for $\tau(\sigma_{V \setminus S_j, S_j} \sigma_{S_j}^{-1}) \sigma_{S_j}$. This implies that if $X \sim W_{Q_G}(\alpha, \beta, \sigma)$, then

$$E_{W_{Q_G}(\alpha, \beta, \sigma)}(X) = \left(\sum_{j=1}^k \alpha_j - \sum_{j=2}^k \beta_j \right) \hat{\sigma} + \sum_{j=1}^k \alpha_j \hat{\sigma}_{V \setminus C_j, C_j} - \sum_{j=2}^k \beta_j \hat{\sigma}_{V \setminus S_j, S_j}$$

We have parallel results for the type II Wishart. We note that for $(\alpha, \beta) \in \mathcal{B}$

$$\mathcal{F}_{(\alpha, \beta), II} = \{W_{P_G}(\alpha, \beta, \theta; dy), \theta \in Q_G\}$$

is the natural exponential family generated by the measure

$$\nu_{(\alpha,\beta),G}(dy) = \frac{H_G(\alpha, \beta; \varphi(y))}{\Gamma_{II}(\alpha, \beta)} \nu_G(dy).$$

Here the Laplace transform $L_{\nu_{(\alpha,\beta),G}}(x)$ is finite if and only if x is in $-Q_G$ (Theorem 2.1 part 1.) which is an open subset of I_G (see Letac and Massam, 2006), and the domain of the means of $\mathcal{F}_{\alpha,\beta,II}$ is P_G . We have $L_{\nu_{(\alpha,\beta),G}}(x) = H_G(\alpha, \beta, -x)$. This implies that the Laplace transform of $W_{P_G}(\alpha, \beta; \theta)$ is defined for $-x + \theta \in Q_G$ by

$$\int_{P_G} e^{\langle x, y \rangle} W_{P_G}(\alpha, \beta, \theta; dy) = \frac{H_G(\alpha, \beta; \theta - x)}{H_G(\alpha, \beta; \theta)}. \quad (4.23)$$

The cumulant transform is

$$k_{\nu_{(\alpha,\beta),G}}(x) = \sum_{j=1}^k \alpha_j \log \det((-x)_{C_j}) - \sum_{j=2}^k \beta_j \log \det((-x)_{S_j}).$$

and its differential is given by the following element of P_G

$$k'_\nu(x) = \sum_{j=1}^k \alpha_j (-x_{C_j}^{-1})^0 - \sum_{j=2}^k \beta_j (-x_{S_j}^{-1})^0. \quad (4.24)$$

This implies that if $Y \sim W_{P_G}(\alpha, \beta, \theta)$, then

$$E_{W_{P_G}(\alpha,\beta,\theta)}(Y) = \sum_{j=1}^k \alpha_j (\theta_{C_j}^{-1})^0 - \sum_{j=2}^k \beta_j (\theta_{S_j}^{-1})^0. \quad (4.25)$$

To conclude this section, let us make a few remarks. Formula (4.25) with Y replaced by $\widehat{\Sigma}_G^{-1}$, (α, β) replaced by $(\alpha' - \alpha, \beta' - \beta)$ and θ replaced by $D + x$ gives the posterior mean $E(\widehat{\Sigma}_G^{-1} | X = x)$ of the inverse of the natural canonical parameter Σ_G for $W_{Q_G}(\alpha, \beta, \Sigma_G)$ in Theorem 4.1 when the prior distribution on Σ_G is the $IW_{P_G}(\alpha', \beta', D)$. It is, of course, also of interest to compute the posterior mean of the natural canonical parameter Σ_G . In other words, we need $E_{W_{P_G}(\alpha' - \alpha, \beta' - \beta, D+x)}(\varphi(Y))$ when $Y = \widehat{\Sigma}_G^{-1}$. We have the general formula

$$\begin{aligned} -\theta &= \left(\sum_{j=1}^k \left(\alpha_j + \frac{c_j + 1}{2} \right) - \sum_{j=2}^k \left(\beta_j + \frac{s_j + 1}{2} \right) \right) E_{W_{P_G}(\alpha,\beta,\theta)}(\varphi(Y)) \\ &\quad + \sum_{j=1}^k \left(\alpha_j + \frac{c_j + 1}{2} \right) E_{W_{P_G}(\alpha,\beta,\theta)}(\varphi(Y)_{V \setminus C_j \cdot C_j}) \\ &\quad - \sum_{j=2}^k \left(\beta_j + \frac{s_j + 1}{2} \right) E_{W_{P_G}(\alpha,\beta,\theta)}(\varphi(Y)_{V \setminus S_j \cdot S_j}). \end{aligned} \quad (4.26)$$

The proof is not straightforward. To derive (4.26), we use Stokes formula and obtain

$$0 = \int_{P_G} (u'(y)v(y) + u(y)v'(y)) dy$$

with $u(y) = H_G(\alpha + \frac{c+1}{2}, \beta + \frac{s+1}{2}, \varphi(y))$, $v(y) = e^{\langle \theta, y \rangle}$ and where $(\alpha, \beta) \in \mathcal{B}$ satisfy some restrictions similar to the restrictions for the existence of the expected value of the inverse Wishart, that is $p > \frac{r+1}{2}$ for the $w_r(p, \sigma)$ distribution. Using the same substitutions for α, β, θ in (4.26) as we did in (4.25), we see that (4.26) implies that in Theorem 4.1, the posterior mean of Σ_G is not linear in x ; this is in accordance with Theorem 3 of Diaconis-Ylvisaker (1979). We also see from Corollary 4.1 that when $X = \pi(S) \in Q_G$ follows the hyper Wishart distribution with $(\alpha, \beta) = (\frac{n}{2}, \frac{n}{2})$ and the shape parameter of the prior on Σ_G are (α', β') , then the shape parameters of the posterior are $(\alpha' - \frac{n}{2}, \beta' - \frac{n}{2})$. That is, as for the inverse Wishart or the hyper inverse Wishart, the parameters (α', β') of the IW_{P_G} are added to half of the sample size and from (4.26) and the choice of (α', β') has the same kind of impact on the posterior mean as the choice of the shape parameters for the inverse or hyper inverse Wishart.

Let us also mention here that the $IW_{P_G}^{**}(\alpha, \beta, \theta)$ distribution as given by equations (4.10)-(4.12) is conditionally $(k+1)$ -reducible and is an enriched conjugate family of prior distributions, in the sense of Consonni and Veronese (2001), for the parameter $\Sigma_G = \pi(\Sigma)$ of a Gaussian distribution Markov with respect to G . This follows immediately from Theorem 4.4. The $IW_{P_G}^{**}(\alpha, \beta, \theta)$ is also closely linked to the enriched standard conjugate Wishart family of priors for $K = \Sigma^{-1}$ in the standard Gaussian distribution, that is when G is a complete graph, built by Consonni and Veronese (2003). Theorem 2 and Corollary 1 in that paper correspond to Theorem 4.4 here. However we should note that it is the $W_{P_G}(\alpha, \beta, \theta)$ family that is an exponential family, not the $IW_{P_G}^{**}(\alpha, \beta, \theta)$ family and therefore the analog of the enriched Wishart, for G decomposable, is the $W_{P_G}(\alpha, \beta, \theta)$.

5 Open problems

We will now raise some natural questions related to the paper.

SINGULARITY. The well known Gyndikin theorem states that the mapping $\theta \mapsto (\det(-\theta))^{-p}$ from $-M_r^+$ to $(0, \infty)$ is the Laplace transform of some positive measure μ_p on symmetric real matrices of order n if and only if p is in the set

$$\Lambda = \left\{ \frac{1}{2}, \frac{2}{2}, \dots, \frac{r-1}{2} \right\} \cup \left(\frac{r-1}{2}, \infty \right).$$

A very readable proof of this theorem can be found in Shanbhag (1988). The natural exponential family generated by μ_p is the set of Wishart distributions with shape parameter p . If $p = j/2$ with $j = 1, \dots, n-1$ then μ_p is concentrated on the singular semipositive definite matrices of rank j . For a decomposable graph G on $V = \{1, \dots, r\}$ and for $(\alpha, \beta) \in \mathcal{A}$, the mapping $y \mapsto H_G(\alpha, \beta; -\varphi(y))$ from $-P_G$ to $(0, \infty)$ is the Laplace transform of a positive

measure on Q_G which generates the natural exponential family of Wishart distributions of type I. Natural questions are

- For which values of α, β is $y \mapsto H_G(\alpha, \beta; -\varphi(y))$ the Laplace transform of some positive measure on I_G ?
- How do we describe these measures?

Similar questions arise with the Wishart distributions of type II: for which values of α, β is the mapping $x \mapsto H_G(\alpha, \beta; -x)$ from $-Q_G$ to $(0, \infty)$ the Laplace transform of some positive measure on Z_G ?

COMPLEX AND QUATERNIONIC NUMBERS. Wishart matrices with complex and quaternionic entries are well defined. Thus many concepts of the present paper are extendable to complex or quaternionic matrices in a rather mechanical way.

THE SETS \mathcal{A} AND \mathcal{B} . Is it true that $\mathcal{A} = \cup A_P$ and that $\mathcal{B} = \cup B_P$ for any non homogeneous graph? Calculations are terrifying for the graph $A_5 : \bullet - \bullet - \bullet - \bullet - \bullet$. On the other hand, this conjecture is easily proved for the tree represented in Figure 3 below, with $n + m + 2$ vertices denoted $a_1, \dots, a_n, b_1, \dots, b_m, 2, 3$ with edges $a_i \sim 2, b_j \sim 3$ for all i, j and $2 \sim 3$: to prove it, we need only extend the calculations of Proposition 3.2 and Corollary 3.1.

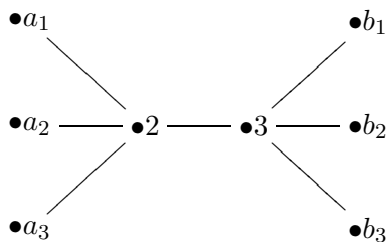


Figure 3: The case $n = m = 3$

Acknowledgements. The authors would like to thank S.A. Andersson for many lively discussions that led to significant improvements in this paper, S.L. Lauritzen for telling the authors that some hyper Markov property must hold for the type I and II Wisharts and finally a referee and the associate editor for reading the manuscript meticulously and suggesting several improvements in the presentation.

6 References

- ABRAMOWITZ, M. AND STEGUN, I. A. (1964). *Handbook of Mathematical Functions*. Dover, New York.
- ANDERSSON, S.A. AND WOJNAR, G. (2004). The Wishart distributions on homogeneous cones. *J. Theoret. Probab.* **17** 781-818.
- BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, New York.
- CONSONNI, G. AND VERONESE, P. (2001). Conditionally reducible natural exponential families and enriched conjugate priors. *Scand. J. Statist.* **28**, 377-406.
- CONSONNI, G. AND VERONESE, P. (2003). Enriched conjugate and reference priors for the Wishart family on the symmetric cones. *Ann. Statist.* **31**, 1491-1516.
- DIACONIS, P. AND YLVIKAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7**, 269-81.
- DAWID, A. AND LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272-1317.
- DEMPSTER, A. (1972). Covariance selection. *Biometrics* **28** 157-175.
- GRÖNE, R., JOHNSON, C. R., SÁ, E. AND WOLKOWICZ, H. (1984). Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.* **58** 109-128.
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.
- LETAC, G. AND MASSAM, H. (2006). Extremal rays and duals for cones of positive definite matrices with prescribed zeros. *Linear Algebra Appl.*, to appear.
- MASSAM, H. AND NEHER, E. (1997). "On transformations and determinants of Wishart variables on symmetric cones," *J. Theoret. Probab.*, **10**, 867-902.
- MASSAM, H. AND NEHER, E. (1998). Estimation and testing for lattice conditional independence models on Euclidean Jordan algebras. *Ann. Statist.* **26** 1051-1082.
- MUIRHEAD, R. (1982). *Aspects of Multivariate Analysis*. Wiley, New York.

OLKIN, I. AND RUBIN, H. (1964). Multivariate Beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist.* **35** 261-269.

ROVERATO, A. (2000). Cholesky decomposition of a hyper Wishart matrix. *Biometrika* **87** 99-112.

SHANBHAG, D. N. (1988). The Davidson Kendall problem and related results on the structure of the Wishart distribution. *Austr. J. Statist.* **30A** 272-280.

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

7 Appendix

7.1 Proofs of §2

Proof of Theorem 2.1 (1). The dual of P_G is $P_G^* = \{x \in I_G; \text{tr}(xy) > 0 \text{ for all } y \in \overline{P_G} \setminus \{0\}\}$ where $\overline{P_G}$ is the closure of P_G , that is, the cone of semi positive definite matrices of Z_G . To show that $Q_G = P_G^*$ we will first show that $Q_G \subset P_G^*$ and then that $Q_G \supset P_G^*$. If $x \in Q_G$ then by Theorem 2.1 there exists a symmetric positive definite matrix \hat{x} which is the completion of x . Thus $\text{tr}(xy) = \text{tr}(\hat{x}y)$ for all $y \in Z_G$. Furthermore, if $y \in \overline{P_G} \setminus \{0\}$ then $\text{tr}(\hat{x}y) = \text{tr}((\hat{x})^{1/2}y(\hat{x})^{1/2})$. Since the matrix $(\hat{x})^{1/2}y(\hat{x})^{1/2}$ is semi positive definite and non zero, its trace is positive. Thus $x \in P_G^*$ and $Q_G \subset P_G^*$ is proved.

Conversely, take $x \in I_G$ such that $x \in P_G^*$. Fix a clique C and consider a vector v of \mathbb{R}^r such that the components of v which are not in C are 0. Denote by v_C and by x_C the restrictions of v to C and to $C \times C$ respectively and assume that $v \neq 0$ and thus $v_C \neq 0$. Since vv^t is in $\overline{P_G} \setminus \{0\}$ and since $x \in P_G^*$, $0 < \text{tr}(xvv^t) = v^t x v = v_C^t x_C v_C$. Moreover, this is true for any $v_C \neq 0$ and therefore x_C is positive definite. Since this is true for all cliques, we deduce that x is in Q_G and $Q_G \supset P_G^*$ is also proved. We thank S. Andersson for this result and this proof. Our former proof was longer and was relying on the description of the extremal lines of the cones $\overline{P_G}$ and $\overline{Q_G}$ given in Letac and Massam (2006). \square

7.2 Proofs of §3.3

Proof of Proposition 3.1. For convenience, we will agree to write t for a vertex of T corresponding to a clique C while we write q for a vertex of T corresponding to a separator S . As defined in (3.10), $m_v = \sum_{u \preceq v} n_u$ and for any $v \in T$, we have

$$|x_C| = \prod_{u \preceq t} |x_{[u]}|, \quad |x_S| = \prod_{u \preceq q} |x_{[u]}|, \quad |C| = c = m_t, \quad |S| = s = m_q.$$

This means that $|x_{[u]}|$ will appear in $|x_C|$ for any C such that $u \preceq t$ and in any $|x_S|$ for any S such that $u \preceq q$. Therefore

$$\begin{aligned}
& \frac{\prod |x_C|^{\alpha_t - \frac{c+1}{2}}}{\prod |x_S|^{\nu(q)(\beta_q - \frac{s+1}{2})}} \tag{7.1} \\
&= \prod_{u \in T} |x_{[u]}|^{\sum_{u \preceq t} (\alpha_t - \frac{m_t+1}{2}) - \sum_{u \preceq q} \nu(q)(\beta_q - \frac{m_q+1}{2})} \\
&= \prod_{u \in T} |x_{[u]}|^{\left(\sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q)\beta_q\right) |x_{[u]}|^{-\left(\sum_{u \preceq t} \frac{m_t+1}{2} - \sum_{u \preceq q} \nu(q)\frac{m_q+1}{2}\right)}} \\
&= \prod_{u \in T} |x_{[u]}|^{\left(\sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q)\beta_q\right) |x_{[u]}|^{-\left(\sum_{u \prec t} \frac{n_t}{2} + \frac{m_u+1}{2}\right)}} \\
&= \prod_{u \in T} |x_{[u]}|^{\left(\sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q)\beta_q\right) |x_{[u]}|^{-\left(\sum_{u \prec v} \frac{n_v}{2} + \sum_{v \prec u} \frac{n_v}{2} + \frac{n_u+1}{2}\right)}}
\end{aligned}$$

where the third equality above follows from the definition of m_v and the fact that for any vertex q of the tree, $\nu(q)$ is equal to the number of children of q minus 1 (see Part 2 of Proposition 2.2). We now make the change of variables (3.14). The Jacobian of this change of variables is

$$J = \prod_{v \in T} |x_{\langle v \rangle}|^{n_v} = \prod_{v \in T} \left(\prod_{u \prec v} |x_{[u]}| \right)^{n_v} = \prod_{u \in T} |x_{[u]}|^{\sum_{u \prec v} n_v} \tag{7.2}$$

Therefore we obtain the image of $H_G(\alpha, \beta, x)\mu_G(dx)$ as

$$\begin{aligned}
& H_G^*(\alpha, \beta, x)\mu_G^*\left(\prod_{u \in T} dx_{[u]} \cdot d(x_{[u]} x_{\langle u \rangle}^{-1})\right) \tag{7.3} \\
&= \prod_{t \in T} |x_{[t]}|^{\alpha_t - \frac{m_t+1}{2}} \prod_{u \in T, u \neq t} |x_{[u]}|^{\left(\sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q)\beta_q + \sum_{u \prec v} \frac{n_v}{2} - \sum_{v \prec u} \frac{n_v}{2}\right) - \frac{n_u+1}{2}} \\
& \hspace{25em} dx_{[u]} \cdot dx_{[u]} x_{\langle u \rangle}^{-1} \\
&= \prod_{u \in T} |x_{[u]}|^{\left(\sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q)\beta_q + \sum_{u \prec v} \frac{n_v}{2} - \sum_{v \prec u} \frac{n_v}{2}\right) - \frac{n_u+1}{2}} dx_{[u]} \cdot dx_{[u]} x_{\langle u \rangle}^{-1} \\
&= \prod_{u \in T} |x_{[u]}|^{\lambda_u - \frac{n_u+1}{2}} dx_{[u]} \cdot dx_{[u]} x_{\langle u \rangle}^{-1}
\end{aligned}$$

where

$$\lambda_u = \sum_{u \preceq t} \alpha_t - \sum_{u \preceq q} \nu(q)\beta_q + \sum_{u \prec v} \frac{n_v}{2} - \sum_{v \prec u} \frac{n_v}{2}.$$

□

7.3 Proofs of §3.4

In the sequel, in order to avoid numbering difficulties for separators with multiplicity greater than one, we sometimes use the generic notation S for a separator and $\nu(S)$ for its multi-

plicity. However, when it is important to list the separators as they appear from a perfect order of the cliques, we denote the separator S_j by $\langle j \rangle$. The double notation should not cause any difficulty.

Proof of Theorem 3.3. To avoid any ambiguity in the notation in the proof below and all the other proofs in the reminder of the paper, let us recall that the set \mathcal{S} of distinct separators contains $k' \leq k - 1$ elements. For convenience, let us write A for the left hand side of (3.19) and $y = \hat{\sigma}^{-1}$ for $\sigma \in Q_G$. Using the Jacobian (2.13) and formula (2.9), we then have

$$\begin{aligned}
A &= \int_{Q_G} e^{-\langle x, \sigma^{-1} \rangle} \frac{\prod_{j=1}^k |x_{C_j}|^{\alpha_j - \frac{c_j+1}{2}}}{\prod_{S \in \mathcal{S}} |x_S|^{\nu(S)(\beta(S) - \frac{|S|+1}{2})}} dx \\
&= \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}, \sigma_{C_1}^{-1} \rangle} \\
&\quad \prod_{j=2}^k |x_{[j]}|^{\alpha_j - \frac{c_j+1}{2}} e^{-\langle x_{[j]}, \sigma_{[j]}^{-1} \rangle} \prod_{j=2}^k e^{-\langle (x_{[j]} x_{\langle j \rangle}^{-1} - \sigma_{[j]} \sigma_{\langle j \rangle}^{-1}), \sigma_{[j]}^{-1} (x_{[j]} x_{\langle j \rangle}^{-1} - \sigma_{[j]} \sigma_{\langle j \rangle}^{-1}) x_{\langle j \rangle} \rangle} \\
&\quad \prod_{S \in \mathcal{S}} |x_S|^{\sum_{i \in J(P, S)} (\alpha_i - \frac{c_i+1}{2}) - \nu(S)(\beta(S) - \frac{|S|+1}{2})} \prod_{S \in \mathcal{S}} |x_S|^{\sum_{i \in J(P, S)} c_i - \nu(S)|S|} dx_{C_1} \prod_{j=2}^k d(x_{[j]} x_{\langle j \rangle}^{-1}) dx_{[j]}.
\end{aligned}$$

Now, since the cardinality of $J(P, S)$ is equal to $\nu(S)$ and, by assumption 1 of the theorem, all $|x_{S_j}| = |x_{\langle j \rangle}|, j \neq 2$ appear with exponent equal to $\frac{c_j - s_j}{2}$ while $|x_{S_2}| = |x_{\langle 2 \rangle}|$ appears with exponent equal to $\sum_{i \in J(P, S_2)} \alpha_i - \nu(q)\beta_q + \sum_{i \in J(P, S_2)} \frac{c_i - s_2}{2}$ we have

$$\begin{aligned}
A &= \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}, \sigma_{C_1}^{-1} \rangle} |x_{\langle 2 \rangle}|^{\delta_2} \\
&\quad \prod_{j=2}^k |x_{\langle j \rangle}|^{\frac{c_j - s_j}{2}} e^{-\langle (x_{[j]} x_{\langle j \rangle}^{-1} - \sigma_{[j]} \sigma_{\langle j \rangle}^{-1}), \sigma_{[j]}^{-1} (x_{[j]} x_{\langle j \rangle}^{-1} - \sigma_{[j]} \sigma_{\langle j \rangle}^{-1}) x_{\langle j \rangle} \rangle} \\
&\quad |x_{[j]}|^{\alpha_j - \frac{s_j}{2} - \frac{c_j - s_j + 1}{2}} e^{-\langle x_{[j]}, \sigma_{[j]}^{-1} \rangle} dx_{C_1} \prod_{j=2}^k d(x_{[j]} x_{\langle j \rangle}^{-1}) dx_{[j]}.
\end{aligned} \tag{7.4}$$

Clearly $(x_{[k]}, x_{[k]} x_{\langle k \rangle}^{-1})$ is independent of $x_{C_1} \cup \dots \cup C_{k-1} = (x_{C_1}, x_{[j]} x_{\langle j \rangle}^{-1}, x_{[j]}, j = 2, \dots, k-1)$ and $x_{[k]}$ is independent of $x_{[k]} x_{\langle k \rangle}^{-1}$. Therefore holding all other variables fixed, we first integrate with respect to $(x_{[k]}, x_{[k]} x_{\langle k \rangle}^{-1})$. Since $\langle k \rangle \subset C_1 \cup \dots, C_{k-1}$, then $x_{\langle k \rangle}$ is fixed and by Lemma 2.4 we obtain

$$\begin{aligned}
&\int_{M_{c_k - s_k}^+} |x_{[k]}|^{\alpha_k - \frac{s_k}{2} - \frac{c_k - s_k + 1}{2}} e^{-\langle x_{[k]}, \sigma_{[k]}^{-1} \rangle} dx_{[k]}, \\
&\times \int_{L(R^{s_k}, R^{c_k - s_k})} |x_{\langle k \rangle}|^{\frac{c_k - s_k}{2}} e^{-\langle (x_{[k]} x_{\langle k \rangle}^{-1} - \sigma_{[k]} \sigma_{\langle k \rangle}^{-1}), \sigma_{[k]}^{-1} (x_{[k]} x_{\langle k \rangle}^{-1} - \sigma_{[k]} \sigma_{\langle k \rangle}^{-1}) x_{\langle k \rangle} \rangle} d(x_{[k]} x_{\langle k \rangle}^{-1}) \\
&= \Gamma_{c_k - s_k}(\alpha_k - \frac{s_k}{2}) |\sigma_{[k]}|^{\alpha_k} \\
&\times \int_{L(R^{s_k}, R^{c_k - s_k})} |\sigma_{[k]}|^{-\frac{s_k}{2}} |x_{\langle k \rangle}|^{\frac{c_k - s_k}{2}} e^{-\langle (x_{[k]} x_{\langle k \rangle}^{-1} - \sigma_{[k]} \sigma_{\langle k \rangle}^{-1}), \sigma_{[k]}^{-1} (x_{[k]} x_{\langle k \rangle}^{-1} - \sigma_{[k]} \sigma_{\langle k \rangle}^{-1}) x_{\langle k \rangle} \rangle} d(x_{[k]} x_{\langle k \rangle}^{-1}).
\end{aligned}$$

$$= \pi^{\frac{s_k(c_k-s_k)}{2}} \Gamma_{c_k-s_k}(\alpha_k - \frac{s_k}{2}) |\sigma_{[k]}|^{\alpha_k}$$

Repeating this process successively for $j = k-1, \dots, 2$, we obtain

$$A = \prod_{j=2}^k \pi^{\frac{s_j(c_j-s_j)}{2}} \Gamma_{c_j-s_j}(\alpha_j - \frac{s_j}{2}) |\sigma_{[j]}|^{\alpha_j} \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}, \sigma_{C_1}^{-1} \rangle} |x_{\langle 2 \rangle}|^{\delta_2} dx_{C_1}.$$

In this last integral, setting, as in §4.2,

$$x_{[1]} = x_{C_1 \setminus S_2} - x_{C_1 \setminus S_2, S_2} x_{S_2, C_1 \setminus S_2}^{-1}, \quad x_{[12]} = x_{C_1 \setminus S_2, S_2} x_{S_2}^{-1},$$

we make the change of variable

$$x_{C_1} \mapsto (x_{[1]}, x_{[12]}, x_{\langle 2 \rangle}) \quad (7.5)$$

with Jacobian equal to $|x_{\langle 2 \rangle}|^{c_1-s_2}$. Then, by (2.11)

$$\begin{aligned} & \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}, \sigma_{C_1}^{-1} \rangle} |x_{\langle 2 \rangle}|^{\delta_2} dx_{C_1} \\ &= \int |x_{[1]}|^{\alpha_1 - \frac{s_2}{2} - \frac{c_1-s_2+1}{2}} e^{-\langle x_{[1]}, \sigma_{[1]}^{-1} \rangle} dx_{[1]} \quad (7.6) \\ & \times \int_{M_{S_2}^+} |x_{\langle 2 \rangle}|^{\alpha_1 + \frac{c_1-s_2}{2} + \delta_2 - \frac{s_2+1}{2}} e^{-\langle x_{\langle 2 \rangle}, \sigma_{S_2}^{-1} \rangle} \\ & \times \left(\int_{L(R^{s_2}, R^{c_1-s_2})} e^{-\langle (x_{[12]} - \sigma_{[12]}), \sigma_{[1]}^{-1}(x_{[1,2]} - \sigma_{[1,2]})x_{\langle 2 \rangle} \rangle} dx_{[12]} \right) dx_{\langle 2 \rangle} \\ &= \Gamma_{c_1-s_2}(\alpha_1 - \frac{s_2}{2}) |\sigma_{[1]}|^{\alpha_1} \pi^{\frac{s_2(c_1-s_2)}{2}} \int_{M_{S_2}^+} |x_{\langle 2 \rangle}|^{\alpha_1 - \frac{s_2}{2} - \frac{c_1-s_2+1}{2} + c_1-s_2 - \frac{c_1-s_2}{2} + \delta_2} e^{-\langle x_{\langle 2 \rangle}, \sigma_{S_2}^{-1} \rangle} dx_{\langle 2 \rangle} \\ &= \pi^{\frac{s_2(c_1-s_2)}{2}} \Gamma_{c_1-s_2}(\alpha_1 - \frac{s_2}{2}) |\sigma_{[1]}|^{\alpha_1} \int_{M_{S_2}^+} |x_{\langle 2 \rangle}|^{\alpha_1 - \frac{s_2+1}{2} + \delta_2} e^{-\langle x_{\langle 2 \rangle}, \sigma_{S_2}^{-1} \rangle} dx_{\langle 2 \rangle} \\ &= \pi^{\frac{s_2(c_1-s_2)}{2}} \Gamma_{c_1-s_2}(\alpha_1 - \frac{s_2}{2}) |\sigma_{[1]}|^{\alpha_1} |\sigma_{S_2}|^{\alpha_1} |\sigma_{S_2}|^{\delta_2} \Gamma_{s_2}(\alpha_1 + \delta_2) \\ &= \pi^{\frac{s_2(c_1-s_2)}{2}} |\sigma_{C_1}|^{\alpha_1} |\sigma_{S_2}|^{\delta_2} \Gamma_{c_1-s_2}(\alpha_1 - \frac{s_2}{2}) \Gamma_{s_2}(\alpha_1 + \delta_2) \quad (7.7) \end{aligned}$$

Now, let us observe that using the multiplicity of S_2 and Assumption 1 of the theorem, we obtain

$$\prod_{j=2}^k |\sigma_{[j]}|^{\alpha_j} = \frac{|\sigma_{C_2}|^{\alpha_2}}{|\sigma_{S_2}|^{\alpha_2}} \prod_{j=3}^k \frac{|\sigma_{C_j}|^{\alpha_j}}{|\sigma_{S_j}|^{\alpha_j}} = \frac{|\sigma_{C_2}|^{\alpha_2}}{|\sigma_{S_2}|^{\sum_{i \in J(P, S_2)} \alpha_i}} \frac{\prod_{j=3}^k |\sigma_{C_j}|^{\alpha_j}}{\prod_{S \in \mathcal{S}, S \not\supset S_2} |\sigma_S|^{\nu(S)\beta(S)}} \quad (7.8)$$

Combining (7.5), (7.7) and (7.8), we obtain (3.19) with

$$\Gamma_I(\alpha, \beta) = \Gamma_{s_2}(\alpha_1 + \delta_2) \times \pi^{\frac{1}{2}(c_1-s_2)s_2} \Gamma_{c_1-s_2}(\alpha_1 - \frac{s_2}{2}) \prod_{j=2}^k \pi^{\sum_{j=2}^k \frac{1}{2}(c_j-s_j)s_j} \Gamma_{c_j-s_j}(\alpha_j - \frac{s_j}{2}).$$

To obtain (3.20), we use (2.16). \square

Proof of Theorem 3.4. For convenience let us denote by B the left hand side of (3.22). Using first $\varphi(y) = x \in Q_G$ and Jacobian (2.4) and then, making the change of variable

(2.12) with Jacobian (2.13) and using (2.9), we have

$$\begin{aligned}
B &= \int_{Q_G} e^{-\langle \theta, \hat{x}^{-1} \rangle} \frac{\prod_{j=1}^k |x_{C_j}|^{\alpha_j - \frac{c_j+1}{2}}}{\prod_{S \in \mathcal{S}} |x_S|^{\nu(S)(\beta(S) - \frac{|S|+1}{2})}} dx \\
&= \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}^{-1}, \theta_{C_1} \rangle} \\
&\quad \prod_{j=2}^k |x_{[j]}|^{\alpha_j - \frac{c_j+1}{2}} e^{-\langle x_{[j]}^{-1}, \theta_{[j]} \cdot \rangle} \prod_{j=2}^k e^{-\langle (x_{[j]} > x_{<j>}^{-1} - \theta_{[j]} > \theta_{<j>}^{-1}), x_{[j]}^{-1} \cdot (x_{[j]} > x_{<j>}^{-1} - \theta_{[j]} > \theta_{<j>}^{-1}) \theta_{<j>} \rangle} \\
&\quad \prod_{S \in \mathcal{S}} |x_S|^{\sum_{i \in J(P,S)} (\alpha_i - \frac{c_i+1}{2}) - \nu(S)(\beta(S) - \frac{|S|+1}{2})} \prod_{S \in \mathcal{S}} |x_S|^{\sum_{i \in J(P,S)} c_i - \nu(S)|S|} dx_{C_1} \prod_{j=2}^k d(x_{[j]} > x_{<j>}^{-1}) dx_{[j]} \cdot
\end{aligned}$$

By Assumption 1 of the theorem, this is equal to

$$\begin{aligned}
B &= \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}^{-1}, \theta_{C_1} \rangle} |x_{<2>}|^{\gamma_2} \\
&\quad \prod_{j=2}^k |x_{[j]}|^{\alpha_j - \frac{c_j+1}{2}} e^{-\langle x_{[j]}^{-1}, \theta_{[j]} \cdot \rangle} e^{-\langle (x_{[j]} > x_{<j>}^{-1} - \theta_{[j]} > \theta_{<j>}^{-1}), x_{[j]}^{-1} \cdot (x_{[j]} > x_{<j>}^{-1} - \theta_{[j]} > \theta_{<j>}^{-1}) \theta_{<j>} \rangle} \\
&\quad dx_{C_1} \prod_{j=2}^k d(x_{[j]} > x_{<j>}^{-1}) dx_{[j]} \cdot
\end{aligned} \tag{7.9}$$

Clearly $x_{C_1}, (x_{[j]} > x_{<j>}^{-1}, x_{[j]} \cdot, j = 2, \dots, k)$ are mutually independent and

$$B = B_1 \times \prod_{j=2}^k B_j,$$

where

$$B_1 = \int |x_{C_1}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{C_1}^{-1}, \theta_{C_1} \rangle} |x_{<2>}|^{\gamma_2} dx_{C_1}$$

and, using Lemma 2.1 in the third equality below, we have

$$\begin{aligned}
B_j &= \int_{M_{c_j-s_j}^+} \left(\int_{L(R^{s_j}, R^{c_j-s_j})} e^{-\langle (x_{[j]} > x_{<j>}^{-1} - \theta_{[j]} > \theta_{<j>}^{-1}), x_{[j]}^{-1} \cdot (x_{[j]} > x_{<j>}^{-1} - \theta_{[j]} > \theta_{<j>}^{-1}) \theta_{<j>} \rangle} d(x_{[j]} > x_{<j>}^{-1}) \right) \\
&\quad |x_{[j]}|^{\alpha_j - \frac{c_j+1}{2}} e^{-\langle x_{[j]}^{-1}, \theta_{[j]} \cdot \rangle} dx_{[j]}. \\
&= \pi^{\frac{s_j(c_j-s_j)}{2}} \int_{M_{c_j-s_j}^+} |\theta_{<j>}|^{-\frac{c_j-s_j}{2}} |x_{[j]}|^{\alpha_j - \frac{c_j+1}{2} + \frac{s_j}{2}} e^{-\langle x_{[j]}^{-1}, \theta_{[j]} \cdot \rangle} dx_{[j]}. \\
&= \pi^{\frac{s_j(c_j-s_j)}{2}} \int_{M_{c_j-s_j}^+} |\theta_{<j>}|^{-\frac{c_j-s_j}{2}} |x_{[j]}^{-1}|^{-\alpha_j + \frac{c_j-s_j+1}{2} - (c_j-s_j+1)} e^{-\langle x_{[j]}^{-1}, \theta_{[j]} \cdot \rangle} d(x_{[j]}^{-1}) \\
&= \pi^{\frac{s_j(c_j-s_j)}{2}} |\theta_{<j>}|^{-\frac{c_j-s_j}{2}} |\theta_{[j]} \cdot|^{\alpha_j} \Gamma_{c_j-s_j}(-\alpha_j) \cdot
\end{aligned}$$

Therefore

$$B = \prod_{j=2}^k \pi^{\frac{s_j(c_j-s_j)}{2}} |\theta_{<j>}|^{-\frac{c_j-s_j}{2}} |\theta_{[j]} \cdot|^{\alpha_j} \Gamma_{c_j-s_j}(-\alpha_j) \times B_1. \tag{7.10}$$

To compute B_1 , let us make the change of variable (7.5). Then

$$B_1 = \int |x_{[1]}|^{\alpha_1 - \frac{c_1+1}{2}} e^{-\langle x_{[1]}^{-1}, \theta_{[1]} \rangle} e^{-\langle (x_{[12]} - \theta_{[12]}), \mathbf{x}_{[1]}^{-1} \cdot (x_{[12]} - \theta_{[12]}) \theta_{\langle 2 \rangle} \rangle} e^{-\langle x_{\langle 2 \rangle}^{-1}, \theta_{\langle 2 \rangle} \rangle} |x_{\langle 2 \rangle}|^{\alpha_1 - \frac{c_1-s_2}{2} - \frac{s_2+1}{2} + \gamma_2 + (c_1-s_2)} dx_{[1]} dx_{[12]} dx_{\langle 2 \rangle}. \quad (7.11)$$

Integrating with respect with $x_{[12]}$ and using Lemma 2.1, we obtain

$$\begin{aligned} B_1 &= |\theta_{\langle 2 \rangle}|^{-\frac{c_1-s_2}{2}} \pi^{\frac{(c_1-s_2)s_2}{2}} \int_{M_{c_1-s_2}^+} e^{-\langle x_{[1]}^{-1}, \theta_{[1]} \rangle} |x_{[1]}|^{\alpha_1 - \frac{c_1-s_2+1}{2}} dx_{[1]} \\ &\times \int_{M_{s_2}^+} e^{-\langle x_{\langle 2 \rangle}^{-1}, \theta_{\langle 2 \rangle} \rangle} |x_{\langle 2 \rangle}|^{\alpha_1 - \frac{c_1-s_2}{2} + \gamma_2 - \frac{s_2+1}{2} + (c_1-s_2)} dx_{\langle 2 \rangle} \\ &= |\theta_{\langle 2 \rangle}|^{-\frac{c_1-s_2}{2}} \pi^{\frac{(c_1-s_2)s_2}{2}} \int_{M_{c_1-s_2}^+} e^{-\langle x_{[1]}^{-1}, \theta_{[1]} \rangle} |x_{[1]}^{-1}|^{-\alpha_1 - \frac{c_1-s_2+1}{2}} d(x_{[1]}^{-1}) \\ &\times \int_{M_{s_2}^+} e^{-\langle x_{\langle 2 \rangle}^{-1}, \theta_{\langle 2 \rangle} \rangle} |x_{\langle 2 \rangle}^{-1}|^{-\alpha_1 - \frac{c_1-s_2}{2} - \gamma_2 - \frac{s_2+1}{2}} d(x_{\langle 2 \rangle}^{-1}) \\ &= \pi^{\frac{(c_1-s_2)s_2}{2}} |\theta_{\langle 2 \rangle}|^{-\frac{c_1-s_2}{2}} |\theta_{[1]}|^{\alpha_1} \Gamma_{c_1-s_2}(-\alpha_1) |\theta_{\langle 2 \rangle}|^{\alpha_1 + \frac{c_1-s_2}{2} + \gamma_2} \Gamma_{s_2}(-\alpha_1 - \frac{c_1-s_2}{2} - \gamma_2) \\ &= \pi^{\frac{(c_1-s_2)s_2}{2}} |\theta_{\langle 2 \rangle}|^{\alpha_1 + \gamma_2} |\theta_{[1]}|^{\alpha_1} \Gamma_{c_1-s_2}(-\alpha_1) \Gamma_{s_2}(-\alpha_1 - \frac{c_1-s_2}{2} - \gamma_2) \\ &= \pi^{\frac{(c_1-s_2)s_2}{2}} |\theta_{C_1}|^{\alpha_1} |\theta_{\langle 2 \rangle}|^{\alpha_1} \Gamma_{c_1-s_2}(-\alpha_1) \Gamma_{s_2}(-\alpha_1 - \frac{c_1-s_2}{2} - \gamma_2). \end{aligned} \quad (7.12)$$

Let us now observe that

$$\begin{aligned} \prod_{j=2}^k |\theta_{\langle j \rangle}|^{-\frac{c_j-s_j}{2}} |\theta_{[j]}|^{\alpha_j} &= \frac{\prod_{j=2}^k |\theta_{C_j}|^{\alpha_j} |\theta_{\langle j \rangle}|^{-\alpha_j - \frac{c_j-s_j}{2}} \prod_{S \in \mathcal{S}} |\theta_S|^{\nu(S)\beta(S)}}{\prod_{S \in \mathcal{S}} |\theta_S|^{\nu(S)\beta(S)}} \\ &= \frac{\prod_{j=2}^k |\theta_{C_j}|^{\alpha_j}}{\prod_{S \in \mathcal{S}} |\theta_S|^{\nu(S)\beta(S)}} \prod_{S \in \mathcal{S}} |\theta_S|^{-\sum_{j \in J(P,S)} \alpha_j + \frac{1}{2}(c_j-s_j) + \nu(S)\beta(S)} \\ &= \frac{\prod_{j=2}^k |\theta_{C_j}|^{\alpha_j}}{\prod_{S \in \mathcal{S}} |\theta_S|^{\nu(S)\beta(S)}} |\theta_{\langle 2 \rangle}|^{-\gamma_2} \end{aligned} \quad (7.13)$$

Combining (7.10), (7.12) and (7.13), we obtain

$$\Gamma_{II}(\alpha, \beta) = \pi^{\frac{1}{2}((c_1-s_2)s_2 + \sum_{j=2}^k (c_j-s_j)s_j)} \Gamma_{s_2} \left[-\alpha_1 - \frac{c_1-s_2}{2} - \gamma_2 \right] \Gamma_{c_1-s_2}(-\alpha_1) \prod_{j=2}^k \Gamma_{c_j-s_j}(-\alpha_j)$$

To obtain (3.23), we use (2.17). \square

7.4 Proofs of §4.3

Proposition 7.1 For $-y \in M_r^+$ and $C \subset \{1, \dots, r\}$ denote $\sigma_C(y) = ((-y)^{-1})_C$. We write y by blocks corresponding to C and its complement

$$y = \begin{bmatrix} y_1 & y_{12} \\ y_{21} & y_2 \end{bmatrix}.$$

We denote for simplicity $y'_{12} = y_{12}y_2^{-1}$. With these notations the differential of $y \mapsto \sigma_C(y)$ is

$$h = \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \mapsto \begin{bmatrix} \sigma_C & -\sigma_C y'_{12} \\ -y'_{21} \sigma_C & y'_{21} \sigma_C y'_{12} \end{bmatrix} \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \begin{bmatrix} \sigma_C \\ -y'_{21} \sigma_C \end{bmatrix}. \quad (7.14)$$

Furthermore the differential of $y \mapsto \kappa_C(y) = -\log \det \sigma_C(y)$ est

$$h \mapsto \operatorname{tr} \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \begin{bmatrix} \sigma_C & -\sigma_C y'_{12} \\ -y'_{21} \sigma_C & y'_{21} \sigma_C y'_{12} \end{bmatrix} = \operatorname{tr} \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \begin{bmatrix} \sigma_C & \hat{\sigma}_{C, V \setminus C} \\ \hat{\sigma}_{V \setminus C, C} & \hat{\sigma}_{V \setminus C, C} \sigma_C^{-1} \hat{\sigma}_{C, V \setminus C} \end{bmatrix}, \quad (7.15)$$

where the last equality is due to the fact that $y'_{12} = -\sigma_C^{-1} \hat{\sigma}_{C, V \setminus C}$.

Proof: We know that

$$\sigma_C(y) = (y^{-1})_1 = -(y_1 - y_{12}y_2^{-1}y_{21})^{-1}.$$

Let M_C and M_C^+ denote the restrictions of M and M_r^+ to the clique C . Then $\sigma_C(y) = a \circ b(y)$ where $a : -M_C^+ \rightarrow M_C^+$ is defined by $a(x) = -x^{-1}$ and has differential $h \mapsto a'(x)(h) = x^{-1}hx^{-1}$ (a linear application from M_C to M_C) and where $b : -M^+ \rightarrow -M_C^+$ is defined by $b(y) = y_1 - y_{12}y_2^{-1}y_{21}$. The differential of b is the following linear mapping from M to M_C :

$$\begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \mapsto h_1 - h_{12}y_2^{-1}y_{21} - y_{12}y_2^{-1}h_{21} + y_{12}y_2^{-1}h_2y_2^{-1}y_{21} = \begin{bmatrix} 1 & -y'_{12} \end{bmatrix} \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \begin{bmatrix} 1 \\ -y'_{21} \end{bmatrix} \quad (7.16)$$

Finally we apply the composition of differentials to obtain

$$\sigma'_C(y)(h) = (a \circ b)'(y)(h) = a'(b(y))(b'(y)(h)) = \sigma_C(y)(b'(y)(h))\sigma_C(y)$$

which gives (7.14) when combined with (7.16). Now consider the real function l defined on M_C^+ by $l(x) = \log \det x$. Then its differential is the linear form on M_C defined by $h \mapsto l'(x)(h) = \operatorname{tr}(x^{-1}h)$. Thus the differential of the real function on M_r^+ defined by $l \circ \sigma_C$ is the following linear form on M

$$h \mapsto (l \circ \sigma_C)'(y)(h) = \operatorname{tr} \sigma_C^{-1} \sigma'_C(y)(h)$$

which gives (7.15) when combined with (7.14). \square

We will now use the previous proposition to compute $m = k'_\mu(y)$. We need to introduce the notation h_C for the restriction of $h \in Z_G$ to $C \times C$ when $C \subset V$ the notation $C' = V \setminus C$ and the notations $h_{C, C'}$ and $h_{C', C}$ for the restrictions of h to $C \times C'$ and $C' \times C$ respectively.

Proposition 7.2 *The differential of the real function $y \mapsto k_\mu(y)$ defined on P_G by (4.20) is the linear form on Z_G defined by*

$$\begin{aligned}
h &\mapsto \sum_{j=1}^k \alpha_j [\operatorname{tr}(h_{C_j} \sigma_{C_j}) - 2 \operatorname{tr}(h_{C'_j, C_j} \sigma_{C_j} y'_{C_j, C'_j}) + \operatorname{tr}(h_{C'_j} y'_{C'_j, C_j} \sigma_{C_j} y'_{C_j, C'_j})] \\
&\quad - \sum_{j=2}^k \beta_j [\operatorname{tr}(h_{S_j} \sigma_{S_j}) - 2 \operatorname{tr}(h_{S'_j, S_j} \sigma_{S_j} y'_{S_j, S'_j}) + \operatorname{tr}(h_{S'_j} y'_{S'_j, S_j} \sigma_{S_j} y'_{S_j, S'_j})] \\
&= \operatorname{tr} \begin{pmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{pmatrix} \left[\sum_{j=1}^k \alpha_j \begin{pmatrix} \sigma_{C_j} & \hat{\sigma}_{C_j, V \setminus C_j} \\ \hat{\sigma}_{V \setminus C_j, C_j} & \hat{\sigma}_{V \setminus C_j, C_j} \sigma_{C_j}^{-1} \hat{\sigma}_{C_j, V \setminus C_j} \end{pmatrix} \right. \\
&\quad \left. - \sum_{j=2}^k \beta_j \begin{pmatrix} \sigma_{S_j} & \hat{\sigma}_{S_j, V \setminus S_j} \\ \hat{\sigma}_{V \setminus S_j, S_j} & \hat{\sigma}_{V \setminus S_j, S_j} \sigma_{S_j}^{-1} \hat{\sigma}_{S_j, V \setminus S_j} \end{pmatrix} \right]
\end{aligned}$$

Proof: We apply (7.15) to each term of the sum k_μ . The proposition has been established for the cone M_r^+ and we apply it here to the restriction $P_G = M_r^+ \cap Z_G$ of M_r^+ . Therefore the formulas for the differentials of functions restricted to this subspace are still in force when interpreted as linear applications defined on Z_G . \square