Universität Augsburg

Institut für
Mathematik

Steen A. Andersson, Thomas Klein

On Riesz and Wishart Distributions Associated with Decomposable Undirected Graphs

Preprint Nr. 06/2008 — 06. Februar 2008
Institut für Mathematik, Universitätsstraße, D-86 135 Augsburg http://www.math.uni-augsburg.de/

Impressum:

Herausgeber:

Institut für Mathematik
Universität Augsburg
86135 Augsburg
`http://www.math.uni-augsburg.de/forschung/preprint/`

ViSdP:

Thomas Klein
Institut für Mathematik
Universität Augsburg
86135 Augsburg

ON RIESZ AND WISHART DISTRIBUTIONS ASSOCIATED WITH DECOMPOSABLE UNDIRECTED GRAPHS

STEEN A. ANDERSSON[◇] AND THOMAS KLEIN
INDIANA UNIVERSITY AND UNIVERSITÄT AUGSBURG

Abstract. Classical Wishart distributions on the open convex cones of positive definite matrices and their fundamental features are extended to generalized Riesz and Wishart distributions associated with decomposable undirected graphs using the basic theory of exponential families. The families of these distributions are parameterized by their expectations/natural parameter and multivariate shape parameter and have a non-trivial overlap with the generalized Wishart distributions defined in Andersson & Wojnar (2004a) and Andersson & Wojnar (2004b). This work also extends the Wishart distributions of type I in Letac & Massam (2007) and, more importantly, presents an alternative point of view on the latter paper.

1. Introduction.

The classical Wishart distribution arises as the distribution of the maximum likelihood (ML) estimator $\hat{\Sigma}$ of the unknown covariance matrix $\Sigma \in \mathbf{PD}(V)$ from a sample of N observables from a multivariate centered normal distribution on $\mathbb{R}^V$, where V is a finite set[1] and $\mathbf{PD}(V)$ denotes the open convex cone of $V \times V$ positive definite matrices. The ML estimator exists[2] with probability one if and only if $N \geq V$. The distribution of the ML estimator is the classical Wishart distribution with *multivariate scale* $\frac{1}{N}\Sigma$ and $f \equiv N$ *degrees of freedom*. This distribution was first derived by Wishart (1928). In the present work it is more convenient to parameterize the Wishart distributions by their expectations and *shape parameter* $\lambda := \frac{f}{2} \geq \frac{V}{2}$. The Wishart distribution with expectation Σ and shape parameter λ is denoted by $\mathbb{W}_{\Sigma,\lambda}$ and is thus given by (1) below. In fact, definition (1) is meaningful for any $\lambda \in \left]\frac{V-1}{2}, \infty\right[$. In the case $V = 1$, this extension of the range of possible values of the shape parameter reduces to the well-known inclusion of the family of χ^2-distributions with integer degrees of freedom and positive scale into the family of gamma distributions. Using Laplace transforms or characteristic functions one may also define Wishart distributions for the shape parameter values $\lambda = \frac{V-1}{2}, \frac{V-2}{2}, \cdots, \frac{1}{2}, 0$. These distributions, called *singular Wishart distributions*, have no density with respect to a Lebesgue measure, are concentrated on certain sets of positive semidefinite $V \times V$ matrices of rank less than V, and are beyond

Date: January 31, 2008
2000 Mathematics Subject Classification. Primary: 62H99; Secondary: 62E99, 15A30.
Key words and phrases. Decomposable undirected graphs, exponential families, graphical models, Riesz distribution, Siegel integral, Wishart distribution.
[◇]Research supported by U.S. National Security Agency Grant number H98230-06-1-0054.

[1]The cardinality of a finite set I is also denoted by I. The context always precludes misunderstanding.

[2]In the opposite case $N < V$ the ML estimator does not exist for any observation.

the scope of this paper, cf. Casalis & Letac (1996), Letac & Massam (1998), and their references.

The family of classical Wishart distributions on the sample space $\mathbf{PD}(V)$ with fixed shape parameter $\lambda \in \left]\frac{V-1}{2}, \infty\right[$ constitutes a statistical model[1] in its own right,

$$\left(\mathbb{W}_{\Sigma,\lambda} \in \mathcal{P}(\mathbf{PD}(V)) \,\middle|\, \Sigma \in \mathbf{PD}(V)\right),$$

called the *classical Wishart model*. We emphasize that the shape parameter is considered to be known and that the sample space and parameter set are identical. Any subset $\mathbf{P} \subset \mathbf{PD}(V)$ represents what we usually call inference in the covariance structure. Since the ML estimator $\hat{\Sigma}$ mentioned above is sufficient and complete further inference in the covariance structure is usually performed in the Wishart model and is meaningful even without any reference to the sample from a multivariate normal distribution. The subset $\mathbf{P}$ can be given by, for example, symmetry and/or conditional independence restrictions, cf. the introduction in Andersson & Wojnar (2004b)[2] and the many references therein. Also in this reference the open convex cone $\mathbf{PD}(V)$ is generalized to homogeneous cones, thus giving rise to *general Wishart distributions/models*. These distributions enjoy properties similar to those of classical Wishart distribution, for example convolution properties. The Wishart distributions in Casalis (1991), Casalis & Letac (1996), Massam & Neher (1997), and Letac & Massam (1998) are all defined on symmetric cones, a special case of homogeneous cones.

Subsets $\mathbf{P}$ induced by Markov properties associated with certain graphs, so-called *graphical normal models*, are of particular interest since they are not homogeneous cones in general. In the present paper we shall consider the case of a decomposable undirected graph $\mathcal{U}$ with underlying vertex set V although the corresponding subset $\mathbf{PD}(\mathcal{U}) \subseteq \mathbf{PD}(V)$ might fail to be a convex cone. Nevertheless, it is in linear one-to-one correspondence to an open and convex cone $\mathbf{P}(\mathcal{U})$, cf. Section 3.

In Section 2 we define the so-called *classical Riesz distributions* on $\mathbf{PD}(V)$ which generalize classical Wishart distributions and are based on the classical Riesz integral. Based on a specific representation of $\mathcal{U}$ as an acyclic mixed graph (see Section 4) a fundamental property of the open and convex cone $\mathbf{P}(\mathcal{U})$ is presented in Section 5. In Sections 6 and 7 the fundamental integral, the basis for the definition of generalized Riesz distributions is evaluated. The expectation of a generalized Riesz distribution is derived in Section 10, allowing the reparameterization from the natural parameter to the expectation. Some properties of the family of Riesz distributions, similar to properties of the family of classical Wishart distributions, are verified in Section 11. The present work is compared to a part of the closely related work by Letac & Massam (2007) in Section 12. In Sections 13 and 14 a natural representation of $\mathcal{U}$ is presented and is used to embed the family of Wishart distributions into the family of Riesz distributions in Section 15. Section 16 establishes the present work to coincide with Andersson & Wojnar (2004a) and Andersson & Wojnar (2004b) in the special case when $\mathbf{P}(\mathcal{U})$ is a homogeneous cone.

---

[1]A statistical model is a family $\left(P_\theta \in \mathcal{P}(\Omega) \,\middle|\, \theta \in \Theta\right)$ of probability measures on the same measurable space Ω, called the *observation space*, that is, belonging to the set $\mathcal{P}(\Omega)$ of all probability measures on Ω, parameterized by a set Θ, called the *parameter set*.

[2]In Andersson & Wojnar (2004b) the subset $\mathbf{P}$ is called C.

2. Classical Wishart and Riesz distributions.

Let V be a finite set. Let $\mathbf{S}(V)$ and $\mathbf{PD}(V) \subseteq \mathbf{S}(V)$ denote the vector space of *symmetric* $V \times V$ matrices and the open convex cone of all *positive definite* $V \times V$ matrices, respectively. Let $\Sigma \in \mathbf{PD}(V)$ and $\lambda > \frac{V-1}{2}$. The classical *Wishart distribution* $\mathbb{W}_{\Sigma,\lambda}$ on $\mathbf{PD}(V)$ with *shape parameter* λ and *expectation* Σ is defined by

$$\mathrm{d}\mathbb{W}_{\Sigma,\lambda}(S) := \frac{\pi^{-\frac{V\,(V-1)}{4}}\lambda^{\lambda V}|S|^{\lambda-\frac{V+1}{2}}}{\prod\left(\Gamma(\lambda - \frac{i-1}{2})\middle|\, i = 1, \cdots, V\right)|\Sigma|^{\lambda}} \exp\{-\lambda \operatorname{tr}(\Sigma^{-1}S)\}\,\mathrm{d}S, \tag{1}$$

where $|P|$ denotes the determinant of $P \in \mathbf{PD}(V)$ and $\mathrm{d}S$ denotes the standard Lebesgue measure on $\mathbf{S}(V)$ restricted to $\mathbf{PD}(V)$. The parameter Σ deserves its name due to $\mathbb{E}\,(\mathbb{W}_{\Sigma,\lambda}) = \Sigma$, where $\mathbb{E}\,(.)$ denotes expectation.

The statistical model $\left(\mathbb{W}_{\Sigma,\lambda} \in \mathcal{P}(\mathbf{S}(V))\middle|\, \Sigma \in \mathbf{PD}(V)\right)$ is well-known to be a full regular exponential family[1] in its expectation parameterization. The corresponding *natural parameter* is $\Delta := \lambda\Sigma^{-1} \in \mathbf{PD}(V)$. Setting $\mathrm{W}_{\Delta,\lambda} := \mathbb{W}_{\lambda\Delta^{-1},\lambda}$ we have

$$\mathrm{d}\mathrm{W}_{\Delta,\lambda}(S) = \frac{\pi^{-\frac{V\,(V-1)}{4}}|\Delta|^{\lambda}|S|^{\lambda-\frac{V+1}{2}}}{\prod\left(\Gamma(\lambda - \frac{i-1}{2})\middle|\, i = 1, \cdots, V\right)} \exp\{-\operatorname{tr}(\Delta S)\}\mathrm{d}S \tag{2}$$

and thus obtain the above distribution in its natural parameterization. Note the notational distinction between Wishart distributions in their expectation parameterization or natural parameterization.

Let $v_1, \cdots, v_V$ be an enumeration of V. As it is customary the set V is then identified with its enumeration, that is, v_i is denoted by i, $i = 1, \cdots, V$. Set $\langle i\rangle := \{1, \cdots, i-1\}$, and note $\langle 1\rangle = \varnothing$. For $i = 1, \cdots, V$ let $\Sigma_{[i]}$, $\Sigma_{\langle i\rangle}$, $\Sigma_{[i\rangle}$, and $\Sigma_{\langle i]}$ denote the $\{i\}\times\{i\}$, the $\langle i\rangle\times\langle i\rangle$, the $\{i\}\times\langle i\rangle$, and the $\langle i\rangle\times\{i\}$ submatrices of $\Sigma \in \mathbf{PD}(V)$, respectively, and define $\Sigma_{[i]\bullet} := \Sigma_{ii} - \Sigma_{[i\rangle}\Sigma^{-1}_{\langle i\rangle}\Sigma_{\langle i]} > 0$, with $\Sigma^{-1}_{\langle i\rangle} := (\Sigma_{\langle i\rangle})^{-1}$. The scalars $\Sigma_{[i]\bullet}$ are rational functions of the entries of Σ and can also be defined via the unique decomposition $\Sigma = TDT^t$, where T is a lower triangular matrix[2] with all diagonal elements equal to 1 and $D =: \mathrm{Diag}(\Sigma_{[i]\bullet}|\, i = 1, \cdots, V)$ is a diagonal matrix[3] with positive diagonal elements. Similarly, we have a unique decomposition $\Delta = U^tEU$, $\Delta \in \mathbf{PD}(V)$, with U being lower triangular with all diagonal elements equal to 1 and $E =: \mathrm{Diag}(\Delta_{[i]\circ}|\, i = 1, \cdots, V)$, thus defining the positive scalars $\Delta_{[i]\circ}$, $i = 1, \cdots, V$.

For any $\Delta \in \mathbf{PD}(V)$ and any $\lambda \equiv (\lambda_i|\, i = 1, \cdots, V) \in \mathbb{R}^V$ the well-defined *Siegel integral*

$$\int_{\mathbf{PD}(V)} \prod\left(S_{[i]\bullet}^{\lambda_i - \frac{V+1}{2}}\middle|\, i = 1, \cdots, V\right) \exp\{-\operatorname{tr}(\Delta S)\}\,\mathrm{d}S$$

is finite if and only if

$$\lambda_i > \tfrac{i-1}{2}, \quad i = 1, \cdots, V, \tag{3}$$

and in that case evaluates to

$$\pi^{\frac{V\,(V-1)}{4}} \prod\left(\frac{\Gamma(\lambda_i - \frac{i-1}{2})}{\Delta_{[i]\circ}^{\lambda_i}}\middle|\, i = 1, \cdots, V\right),$$

---

[1] Wishart distributions are viewed as distributions on the vector space $\mathbf{S}(V)$ in accordance with the definition of exponential families.

[2] Note that *triangularity* of a $V \times V$ matrix is defined relative to the given enumeration of V.

[3] An $I \times I$ diagonal matrix with diagonal $(d_i|\, i \in I) \in \mathbb{R}^I$ is denoted by $\mathrm{Diag}(d_i|\, i \in I)$.

where $\Delta_{[i]\circ}^{\alpha} := (\Delta_{[i]\circ})^{\alpha}$, and $S_{[i]\bullet}^{\alpha} := (S_{[i]\bullet})^{\alpha}$, $\alpha \in \mathbb{R}$, $i = 1, \cdots, V$, see Faraut & Korányi (1994, Theorem VII.1.1)[1]. Hence we may, for all $\Delta \in \mathbf{PD}(V)$ and $\lambda \in \times \left(\left] \frac{i-1}{2}, \infty \right[\middle| \, i = 1, \cdots, V \right)$, define the probability measure on $\mathbf{PD}(V)$

$$(4)\quad \mathrm{dR}_{\Delta,\lambda}(S) := \pi^{-\frac{V(V-1)}{4}} \prod \left(\frac{\Delta_{[i]\circ}^{\lambda_i} S_{[i]\bullet}^{\lambda_i - \frac{V+1}{2}}}{\Gamma(\lambda_i - \frac{i-1}{2})} \middle| \, i = 1, \cdots, V \right) \exp\{-\operatorname{tr}(\Delta S)\} \, \mathrm{d}S.$$

Definition 2.1. The probability measure $\mathrm{R}_{\Delta,\lambda}$ is called the *classical Riesz distribution* on $\mathbf{PD}(V)$ with respect to the ordering $1, \cdots, V$ of the elements of V and with *shape parameter* $\lambda \equiv (\lambda_i | \, i = 1, \cdots, V)$ and *natural parameter* Δ.

The family of classical Riesz distributions was first introduced by Hassairi & Lajmi (2001) under the name *Riesz natural exponential family (Riesz NEF)* and was based on a special case of the so-called Riesz measure from Faraut & Korányi (1994, p. 137), going back to Riesz (1949).

Note that $\lambda = (\mu | \, i = 1, \cdots, V)$ yields $\mathrm{R}_{\Delta,\lambda} = \mathrm{W}_{\Delta,\mu}$.

Given the numbering $1, \cdots, V$ of the elements of V we can, for all $\Sigma \in \mathbf{PD}(V)$ and all $\lambda \equiv (\lambda_i | \, i = 1, \cdots, V) \in \mathbb{R}_+^V$, define the λ-*inverse*

$$\Sigma^{-\lambda} := (T^t)^{-1} \operatorname{Diag} \left(\frac{\lambda_i}{\Sigma_{[i]\bullet}} \middle| \, i = 1, \cdots, V \right) T^{-1} \in \mathbf{PD}(V),$$

where $\Sigma = T \operatorname{Diag}(\Sigma_{[i]\bullet} | \, i = 1, \cdots, V) T^t$ with T being lower triangular with 1's in the diagonal. Clearly the mapping $\mathbf{PD}(V) \to \mathbf{PD}(V)$, $\Sigma \mapsto \Sigma^{-\lambda}$, is a bijection. Note that the definition of $\Sigma^{-\lambda}$ depends on a specific ordering of V. Note also

$$(5)\qquad (\Sigma^{-\lambda})_{[i]\circ} = \frac{\lambda_i}{\Sigma_{[i]\bullet}}, \qquad i = 1, \cdots, V.$$

In the case $\lambda = (\mu | \, i = 1, \cdots, V)$, that is, if λ_i does not depend on $i = 1, \cdots, V$, we set $\Sigma^{-\mu} := \Sigma^{-\lambda}$. Then $\Sigma^{-1} = \Sigma^{-(1| \, i=1,\cdots,V)}$ is the standard matrix inverse of Σ, and we have $\Sigma^{-\mu} = \mu \, \Sigma^{-1}$.

By a simple calculation or as a special case of Proposition 10.1 we have $\mathbb{E}\,(R_{\Delta,\lambda}) = \Sigma$ if $\Delta = \Sigma^{-\lambda}$. Replacing the natural parameter $\Delta \in \mathbf{PD}(V)$ with the *expectation parameter* $\Sigma \in \mathbf{PD}(V)$ we set $\mathbb{R}_{\Sigma,\lambda} := \mathrm{R}_{\Sigma^{-\lambda},\lambda}$ and thus obtain

$$(6)\qquad \mathrm{d}\mathbb{R}_{\Sigma,\lambda}(S) = \pi^{-\frac{V(V-1)}{4}} \prod \left(\frac{\lambda_i^{\lambda_i} \, S_{[i]\bullet}^{\lambda_i - \frac{V+1}{2}}}{\Gamma(\lambda_i - \frac{i-1}{2}) \, \Sigma_{[i]\bullet}^{\lambda_i}} \middle| \, i = 1, \cdots, V \right) \times \exp\{-\operatorname{tr}(\Sigma^{-\lambda} S)\} \, \mathrm{d}S,$$

the Riesz distribution $\mathbb{R}_{\Sigma,\lambda}$ parameterized by its expectation Σ and shape parameter $\lambda \equiv (\lambda_i | \, i = 1, \cdots, V)$.

3. Positive definite matrices and decomposable undirected graphs.

Let $\mathcal{U} \equiv (V, F)$ be a decomposable undirected graph (DUG) with vertex set V and edge set $F \subset V \times V$. We refer the reader to Lauritzen (1996, Chapter 2) for some basic concepts in graph theory. For $S \in \mathbf{S}(V)$, let S_{uv} denote the (u, v) entry of S, $u, v \in V$. Define

$$\mathbf{S}(\mathcal{U}) := \left\{ S \in \mathbf{S}(V) \middle| \, S_{uv} = 0 \text{ for all } u, v \in V \text{ with } u \neq v \text{ and } (u, v) \notin F \right\},$$

[1] Faraut & Korányi (1994) define this integral only for $\Delta = 1_V$, the $V \times V$ identity matrix.

a subspace of $\mathbf{S}(V)$, and the *projection mapping* $p_{\mathcal{U}} \equiv p : \mathbf{S}(V) \to \mathbf{S}(\mathcal{U})$ by

$$p(S)_{uv} := \begin{cases} S_{uv} & \text{if } (u,v) \in F \text{ or } u = v \\ 0 & \text{if } (u,v) \notin F \text{ and } u \neq v \end{cases} \qquad \text{for all } S \in \mathbf{S}(V).$$

Set $\mathbf{PD}^0(\mathcal{U}) := \mathbf{S}(\mathcal{U}) \cap \mathbf{PD}(V)$ and $\mathbf{PD}(\mathcal{U}) := \mathbf{PD}^0(\mathcal{U})^{-1}$. It is well-known that a centered[1] normal distribution N_Σ on $\mathbb{R}^V$ with covariance matrix $\Sigma \in \mathbf{PD}(V)$ satisfies the Markov properties[2] given by $\mathcal{U}$ if and only if $\Delta = \Sigma^{-1} \in \mathbf{PD}^0(\mathcal{U})$. Similarly, the *subset* $\mathbf{PD}(\mathcal{U}) \subseteq \mathbf{PD}(V)$ is characterized by the equivalence of $\Sigma \in \mathbf{PD}(\mathcal{U})$ and N_Σ satisfying the Markov properties given by $\mathcal{U}$. We set

$$\mathbf{P}(\mathcal{U}) := \big\{ S \in \mathbf{S}(\mathcal{U}) \,\big|\, \forall\, C \in \mathcal{C} :\ S_C \in \mathbf{PD}(C) \big\},$$

where $\mathcal{C}$ denotes the set of cliques in $\mathcal{U}$, and S_A denotes the $A \times A$ submatrix[3] of $S \in \mathbf{S}(V)$, $A \subseteq V$.

Proposition 3.1. *The mapping*

$$\mathbf{PD}(\mathcal{U}) \to \mathbf{P}(\mathcal{U}), \quad S \mapsto p(S), \tag{7}$$

is a well defined bijection.

The open convex cones $\mathbf{P}(\mathcal{U})$ *and* $\mathbf{PD}^0(\mathcal{U})$ *are dual to each other through the isomorphism (of open convex cones)*

$$\mathbf{P}(\mathcal{U}) \to \big(\mathbf{PD}^0(\mathcal{U})\big)^*, \quad S \mapsto \big(T \mapsto \operatorname{tr}(ST)\big).$$

Writing the inverse mapping of (7) *as* $\mathbf{P}(\mathcal{U}) \to \mathbf{PD}(\mathcal{U})$, $S \mapsto \tilde{S}$, *we have the (generalized) matrix inverse mappings constructed from* (7),

$$\begin{array}{rcl} \mathbf{P}(\mathcal{U}) & \leftrightarrow & \mathbf{PD}^0(\mathcal{U}), \\ S & \mapsto & (\tilde{S})^{-1} =: S^{-1}, \\ p(T^{-1}) & \leftarrow\!\shortmid & T. \end{array} \tag{8}$$

These mappings constitute a one-to-one correspondence.

Proof. The two first results[4] are well known and proved by induction by the number of cliques in $\mathcal{U}$. The last statement follows from the definitions of $\mathbf{P}(\mathcal{U})$ and $\mathbf{PD}^0(\mathcal{U})$. □

Remark 3.1. Note that the convex cones $\mathbf{PD}^0(\mathcal{U})$ and $\mathbf{P}(\mathcal{U})$ are open subsets of the vector space $\mathbf{S}(\mathcal{U})$ with $\mathbf{PD}^0(\mathcal{U}) \subseteq \mathbf{P}(\mathcal{U})$. In particular, the dimensions of both cones are equal to that of $\mathbf{S}(\mathcal{U})$. If $\mathcal{U}$ is complete we have $\mathbf{PD}^0(\mathcal{U}) = \mathbf{P}(\mathcal{U}) = \mathbf{PD}(V)$. The definition of S^{-1} in (8) then coincides with the classical matrix inverse of S, also denoted by S^{-1}.

Remark 3.2. Let $A, B \subseteq V$ with $u - v$ for all $u \in A$, $v \in B$, and let $S \in \mathbf{S}(V)$. Then $S_{A\times B} = p(S)_{A\times B}$. In particular we have $\tilde{S}_{A\times B} = S_{A\times B}$, $S \in \mathbf{P}(\mathcal{U})$, where $\tilde{S}_{A\times B}$ is understood as $(\tilde{S})_{A\times B}$.

---

[1] Since its expectation is irrelevant for a normal distribution's Markov properties we consider only *centered* normal distributions, that is, normal distributions with expectation 0.

[2] The strong, weak, and pairwise Markov properties are equivalent in this case; see Lauritzen (1996, Chapter 3) for an overview of Markov properties given by undirected graphs.

[3] In general, we will write $M_{A\times B}$ for the $A \times B$ submatrix of any $V \times V$ matrix M, $A, B \subseteq V$.

[4] The results and their proofs are due to the first author of the present paper, cf. Letac & Massam (2007, Section A1).

4. Representations of a decomposable undirected graph as an acyclic mixed graph.

Let $\mathcal{V} \equiv (V, F)$ be an acyclic mixed graph (AMG)[1] with vertex set V and edge set $F \subset V \times V$, cf. Andersson, Madigan & Perlman (2001, Section 2) for the basic concepts. A part of the definition of a mixed graph is that $(v, v) \notin F$, $v \in V$. We write $v_1 - v_2$ if $(v_1, v_2) \in F$ and $(v_2, v_1) \in F$ and $v_1 \to v_2$ if $(v_1, v_2) \in F$ and $(v_2, v_1) \notin F$. The adjective *acyclic* means that the graph has no partially directed cycles. For all vertices $v \in V$ the sets of *parents* and *neighbors* of v are

$$\begin{aligned} \mathrm{pa}_{\mathcal{V}}(v) \equiv \mathrm{pa}(v) &:= \big\{ v' \in V \,\big|\, (v', v) \in F \text{ and } (v, v') \notin F \big\}, \\ \mathrm{nb}_{\mathcal{V}}(v) \equiv \mathrm{nb}(v) &:= \big\{ v' \in V \,\big|\, (v', v) \in F \text{ and } (v, v') \in F \big\}, \end{aligned}$$

respectively. The *underlying undirected graph* or *skeleton* of $\mathcal{V}$ is defined as $\mathcal{U}(\mathcal{V}) := (V, F \cup F^{\mathrm{o}})$ with $(v_1, v_2)^{\mathrm{o}} := (v_2, v_1)$, $(v_1, v_2) \in V \times V$.

Writing $v_1 \sim v_2$ if $v_1, v_2 \in V$ are equal or connected by an undirected path in $\mathcal{V}$ defines an equivalence relation $\sim \equiv \sim_{\mathcal{V}}$ on V. Its equivalence classes are called *boxes*[2], and the set of equivalence classes $V/\!\sim$ is called the *box set of* $\mathcal{V}$. Each *box* $B \in V/\!\sim$ is thus a subset of V, denoted by[3] $[B]$, and the subgraph induced by $[B] \subseteq V$ is an undirected graph, the *B-box graph* $\mathcal{B}_B \equiv \mathcal{B} := \big([B], F \cap ([B] \times [B])\big)$. The box set $V/\!\sim$ is equipped with an edge set $F^{\sim} \subset (V/\!\sim) \times (V/\!\sim)$ in a natural way, where $(B, B') \in F^{\sim}$ holds if and only if there are $v \in [B]$ and $v' \in [B']$ such that $(v, v') \in F$ and $B \neq B'$. The *graph of boxes* $\mathcal{V}/\!\sim := (V/\!\sim, F^{\sim})$ is in fact an acyclic directed graph (ADG)[4]. For $(B, B') \in F^{\sim}$ we also write $B \to B'$. A box B is called *maximal* if $B \not\to B'$ for all $B' \in V/\!\sim$.

We will frequently use the following two assumptions on a given AMG:

(A1) All box graphs are complete.
(A2) The AMG has no triplexes[5], cf. Andersson et al. (2001).

If the AMG $\mathcal{V}$ satisfies (A1) and (A2), then the set $\mathrm{pa}(v)$ does not depend on the choice of $v \in [B]$, $B \in V/\!\sim$. Hence we may set $\langle B \rangle := \mathrm{pa}(v)$ in that case. Note that the subsets $\langle B \rangle \subseteq V$, $[B]$, and $[B] \,\dot{\cup}\, \langle B \rangle$ then induce AMGs with complete skeletons. Furthermore conditions (A1) and (A2) jointly imply the skeleton $\mathcal{U}(\mathcal{V})$ to be a DUG. This gives rise to the following definition.

Definition 4.1. If $\mathcal{U}$ is a DUG and $\mathcal{V}$ is an AMG satisfying (A1), (A2), and $\mathcal{U} = \mathcal{U}(\mathcal{V})$, then we call $\mathcal{V}$ a *representation of* $\mathcal{U}$ *(as an AMG)*.

Every DUG $\mathcal{U}$ can be turned into an ADG without immoralities by converting lines to arrows; as a consequence, every DUG $\mathcal{U}$ has a representation as an AMG.

Note finally that condition (A2) implies the AMP Markov properties given by $\mathcal{V}$ to be equivalent to those given by $\mathcal{U}(\mathcal{V})$, cf. Andersson et al. (2001, Theorem 6.1).

---

[1]AMGs are also called *chain graphs* in the statistical literature.

[2]Boxes are also called *chain components* in the literature.

[3]The box $B \in V/\!\sim$ is the same as $[B]$. We write $[B]$ when we want to emphasize the fact that B is a set of vertices.

[4]An AMG $\mathcal{V}$ is called an *acyclic directed graph* if it contains no lines.

[5]A *triplex* is a subgraph $\bullet - \bullet \leftarrow \bullet$ (a *flag*) or $\bullet \to \bullet \leftarrow \bullet$ (an *immorality*) of $\mathcal{V}$ induced by a three-vertex subset of V.

5. The fundamental decompositions

Let $\mathcal{U}$ be a DUG with vertex set V and let $\mathcal{V}$ be a representation of $\mathcal{U}$ as an AMG. For $S \in \mathbf{S}(V)$ and $B \in V/\!\sim$ define $S_{[B]}$, $S_{[B\rangle}$, $S_{\langle B]}$, and $S_{\langle B\rangle}$ as the $[B] \times [B]$, $[B] \times \langle B\rangle$, $\langle B\rangle \times [B]$, and $\langle B\rangle \times \langle B\rangle$ submatrices of S, respectively. Remark 3.2 then yields[1] $(\tilde{S})_{\langle B\rangle} = S_{\langle B\rangle}$, $\tilde{S}_{[B]} = S_{[B]}$, $\tilde{S}_{\langle B]} = S_{\langle B]}$, and $\tilde{S}_{[B\rangle} = S_{[B\rangle}$, $B \in \mathbb{R}^{V/\sim}$, $S \in \mathbf{P}(\mathcal{U})$.

Let $M \in V/\!\sim$ be a maximal box and let $\mathcal{V}_M$ be the AMG induced by the subset $V_M := V \setminus [M]$. Then we have $V_M/\!\sim = (V/\!\sim) \setminus \{M\}$. Furthermore $[B]$, $\langle B\rangle$, $S_{[B]}$, $S_{[B\rangle}$, and $S_{\langle B\rangle}$ remain unchanged when $\mathcal{V}$ is replaced with $\mathcal{V}_M$, $B \in V_M/\!\sim$. The skeleton $\mathcal{U}_M(\mathcal{V}_M) \equiv \mathcal{U}_M$ of $\mathcal{V}_M$ is the same as the subgraph of $\mathcal{U}$ induced by the subset V_M of V.

In the following we will partition $V \times V$ matrices according to the decomposition $V = \dot{\bigcup}\big([B]\big|\, B \in V/\!\sim\big)$. Let $\mathbf{D}^+(\mathcal{V})$ denote the convex cone of all $V \times V$ block-diagonal matrices $E = \mathrm{Diag}(E_B|\, B \in V/\!\sim)$ with $E_B \in \mathbf{PD}([B])$, $B \in V/\!\sim$, and let $\mathbf{T}^1_\ell(\mathcal{V})$ denote the set of all $V \times V$ matrices $U \equiv \big(U_{uv}\big|\,(u,v) \in V \times V\big)$ with the properties

(i) $U_{vv} = 1$, $v \in V$,
(ii) $U_{uv} = 0$ if $u \neq v$ and $(u,v) \notin \dot{\bigcup}\big([B] \times \langle B\rangle\big|\, B \in V/\!\sim\big)$.

If we number the boxes $B_1, \cdots, B_{V/\sim} \in V/\!\sim$ in $\mathcal{V}$ in a *faithful* manner, that is, such that $B_i \to B_j$ implies $i < j$, $i,j = 1, \cdots, V/\!\sim$, then the blocked matrices $U \equiv (U_{B_i \times B_j}|\, i,j = 1, \cdots, V/\!\sim) \in \mathbf{T}^1_\ell(\mathcal{V})$ will appear as lower block-triangular matrices with $U_B = 1_{[B]}$, the $[B] \times [B]$ identity matrix, $B \in V/\!\sim$, and with possible extra single entries $U_{uv} = 0$ under the block diagonal[2]. Note that $U \in \mathbf{T}^1_\ell(\mathcal{V})$ is completely determined by specifying its submatrices $U_{[B]\times\langle B\rangle}$, $B \in V/\!\sim$.

Proposition 5.1. *The mapping*

$$\begin{array}{rcl} \mathbf{T}^1_\ell(\mathcal{V}) \times \mathbf{D}^+(\mathcal{V}) & \to & \mathbf{PD}^0(\mathcal{U}) \\ (U,E) & \mapsto & U^t E U \end{array} \tag{9}$$

is a well-defined bijection.

Proof. We shall use induction by the number of boxes in $V/\!\sim$. Let $M \in V/\!\sim$ be a maximal box. Partitioning all matrices according to the decomposition $V = V_M \,\dot{\cup}\, [M]$ yields

$$\begin{aligned} & U^t E U \\ &= \begin{pmatrix} U^t_{V_M\times V_M} & U^t_{[M]\times V_M} \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} E_{V_M\times V_M} & 0 \\ 0 & E_{[M]} \end{pmatrix} \begin{pmatrix} U_{V_M\times V_M} & 0 \\ U_{[M]\times V_M} & 1_{[M]} \end{pmatrix} \\ &= \begin{pmatrix} U^t_{V_M\times V_M} E_{V_M\times V_M} U_{V_M\times V_M} + U^t_{[M]\times V_M} E_{[M]} U_{[M]\times V_M} & U^t_{[M]\times V_M} E_{[M]} \\ E_{[M]} U_{[M]\times V_M} & E_{[M]} \end{pmatrix}. \end{aligned} \tag{10}$$

Clearly we have $U_{V_M\times V_M} \in \mathbf{T}^1_\ell(\mathcal{V}_M)$ and $E_{V_M\times V_M} \in \mathbf{D}^+(\mathcal{V}_M)$. Hence the first term in the upper left corner belongs to $\mathbf{PD}^0(\mathcal{U}_M)$ by the induction assumption. The second term in the upper left corner is positive semidefinite; hence we only have to verify the (u,v) entry to be zero for all $(u,v) \notin F \cup F^o$. We may assume

[1] The undirected graphs induced by $\langle B\rangle \subset V$ and $[B] \subseteq V$ are complete, and for all $v \in [B]$, $u \in \langle B\rangle$, we have $u \to v$.

[2] *Under the block diagonal* means: $(u,v) \in [B_i] \times [B_j]$ for some $i > j$.

$u \in V_M \setminus \langle M\rangle$ or $v \in V_M \setminus \langle M\rangle$ since $\langle M\rangle$ induces a complete subgraph of $\mathcal{U}$. Thus we have

$$\begin{aligned}(U^t_{[M]\times V_M} E_{[M]} U_{[M]\times V_M})_{uv} &= \sum \big((U^t_{[M]\times V_M})_{ux}(E_{[M]})_{xy}(U_{[M]\times V_M})_{yv}\big|\, x,y \in [M]\big)\\ &= \sum \big(U_{xu}E_{xy}U_{yv}\big|\, x,y\in[M]\big) = 0.\end{aligned}$$

For the lower left corner of (10) we obtain

$$\begin{aligned}(E_{[M]}U_{[M]\times V_M})_{uv} &= \sum \big((E_{[M]})_{ux}(U_{[M]\times V_M})_{xv}\big|\, x\in[M]\big)\\ &= \sum\big(E_{ux}U_{xv}\big|\, x\in[M]\big) = 0\end{aligned}$$

for $(u,v) \in [M]\times(V_M\setminus\langle M\rangle)$. This establishes the mapping to be well-defined.

Let $\Delta \in \mathbf{PD}^0(\mathcal{U})$. We will prove existence and uniqueness of a solution $(U,E) \in \mathbf{T}^1_\ell(\mathcal{V}) \times \mathbf{D}^+(\mathcal{V})$ to the equation $U^tEU = \Delta$. To this end the equations

$$\begin{aligned}U^t_{V_M\times V_M}E_{V_M\times V_M}U_{V_M\times V_M} + U^t_{[M]\times V_M}E_{[M]}U_{[M]\times V_M} &= \Delta_{V_M\times V_M}\,,\\ E_{[M]}U_{[M]\times V_M} &= \Delta_{[M]\times V_M}\,,\\ E_{[M]} &= \Delta_{[M]}\end{aligned}$$

have to be solved for (U,E). Doing so we obtain

$$\begin{aligned}E_{[M]} &= \Delta_{[M]}\,,\\ U_{[M]\times V_M} &= \Delta^{-1}_{[M]}\Delta_{[M]\times V_M}\,,\\ U^t_{V_M\times V_M}E_{V_M\times V_M}U_{V_M\times V_M} &= \Delta_{V_M\times V_M} - \Delta_{V_M\times[M]}\Delta^{-1}_{[M]}\Delta_{[M]\times V_M}\,.\end{aligned}$$

Calculations similar to those above establish

$$\big(\Delta^{-1}_{[M]}\Delta_{[M]\times V_M}\big)_{[M]\times(V_M\setminus\langle M\rangle)} = \Delta^{-1}_{[M]}\Delta_{[M]\times(V_M\setminus\langle M\rangle)} = 0$$

and $\Delta_{V_M\times V_M} - \Delta_{V_M\times[M]}\Delta^{-1}_{[M]}\Delta_{[M]\times V_M} \in \mathbf{PD}^0(\mathcal{U}_M)$. By the induction assumption there is a uniquely determined pair $(U_{V_M\times V_M}, E_{V_M\times V_M}) \in \mathbf{T}^1_\ell(\mathcal{V}_M)\times\mathbf{D}^+(\mathcal{V}_M)$ satisfying the third equation. Thus we have found a unique solution. This establishes the mapping to be a bijection. □

Remark 5.1. Proposition 5.1 could also be obtained using the properties of the normal distribution and a combination of Andersson & Perlman (1998, Section 11) and Andersson et al. (2001, Section 5).

The following corollary is a trivial consequence of Proposition 5.1.

Corollary 5.1. *The mapping* $\mathbf{T}^1_\ell(\mathcal{V})\times\mathbf{D}^+(\mathcal{V}) \to \mathbf{P}(\mathcal{U})$, $(U,D)\mapsto p\big(U^{-1}D(U^t)^{-1}\big)$, *is a bijection.*

For $\Delta \in \mathbf{PD}^0(\mathcal{U})$ we define the matrices $\Delta_{[B]^\circ} \in \mathbf{PD}([B])$, $B \in V/\!\sim$, through $\mathrm{Diag}(\Delta_{[B]^\circ}|\, B \in V/\!\sim) := E$ from the unique solution $(U,E) \in \mathbf{T}^1_\ell(\mathcal{V})\times\mathbf{D}^+(\mathcal{V})$ of $U^tEU = \Delta$.

Corollary 5.2. *Let $M \in V/\sim$ be a maximal box in $\mathcal{V}$. Then the mapping*

$$\mathbf{PD}^0(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M]) \to \mathbf{PD}^0(\mathcal{U}),$$

$$\begin{aligned}(\Delta_M, \Pi_M, \Upsilon_M) &\mapsto \begin{pmatrix} 1_{V_M} & -\Pi_{M0}^t \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} \Delta_M & 0 \\ 0 & \Upsilon_M \end{pmatrix} \begin{pmatrix} 1_{V_M} & 0 \\ -\Pi_{M0} & 1_{[M]} \end{pmatrix} \\ &= \begin{pmatrix} \Delta_M + \Pi_{M0}^t \Upsilon_M \Pi_{M0} & -\Pi_{M0}^t \Upsilon_M \\ -\Upsilon_M \Pi_{M0} & \Upsilon_M \end{pmatrix},\end{aligned} \tag{11}$$

with $\Pi_{M0} \in \mathbb{R}^{[M]\times V_M}$ given by $(\Pi_{M0})_{[M]\times\langle M\rangle} = \Pi_M$ and $(\Pi_{M0})_{[M]\times(V_M\setminus\langle M\rangle)} = 0$ (the $[M] \times (V_M \setminus \langle M\rangle)$ zero matrix), is a well-defined bijection.

Proof. The result follows from the fact that $U \in \mathbf{T}_\ell^1(\mathcal{V})$ may be written as

$$U = \begin{pmatrix} U_{V_M\times V_M} & 0 \\ U_{[M]\times V_M} & 1_{[M]} \end{pmatrix} = \begin{pmatrix} U_{V_M\times V_M} & 0 \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} 1_{V_M} & 0 \\ U_{[M]\times V_M} & 1_{[M]} \end{pmatrix}$$

and by setting $\Pi_{M0} := -U_{[M]\times V_M}$. □

Remark 5.2. Let $\Delta \in \mathbf{PD}^0(\mathcal{U})$ denote the right-hand side of equation (11). Applying Proposition 5.1 to $\Delta_M \in \mathbf{PD}^0(\mathcal{U}_M)$ from (11) we obtain $\Delta_M = U_M^t E_M U_M$ with $(U_M, E_M) \in \mathbf{T}_\ell^1(\mathcal{V}_M) \times \mathbf{D}^+(\mathcal{V}_M)$. Due to

$$\begin{pmatrix} U_M & 0 \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} 1_{V_M} & 0 \\ -\Pi_{M0} & 1_{[M]} \end{pmatrix} = \begin{pmatrix} U_M & 0 \\ -\Pi_{M0} & 1_{[M]} \end{pmatrix} \in \mathbf{T}_\ell^1(\mathcal{V})$$

Proposition 5.1 implies $(\Delta_M)_{[B]\circ} = \Delta_{B\circ}$, $B \in V_M/\sim$.

The following proposition states our fundamental decomposition result for matrices in $\mathbf{P}(\mathcal{U})$.

Proposition 5.2. *Let $M \in V/\sim$ be a maximal box in $\mathcal{V}$. Then the mapping*

$$\mathbf{P}(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M]) \to \mathbf{P}(\mathcal{U}),$$

$$\begin{aligned}(S_M, R_M, L_M) &\mapsto p\left(\begin{pmatrix} 1_{V_M} & 0 \\ R_{M0} & 1_{[M]} \end{pmatrix} \begin{pmatrix} \tilde{S}_M & 0 \\ 0 & L_M \end{pmatrix} \begin{pmatrix} 1_{V_M} & R_{M0}^t \\ 0 & 1_{[M]} \end{pmatrix}\right) \\ &= p\left(\begin{pmatrix} \tilde{S}_M & \tilde{S}_M R_{M0}^t \\ R_{M0}\tilde{S}_M & L_M + R_{M0}\tilde{S}_M R_{M0}^t \end{pmatrix}\right),\end{aligned} \tag{12}$$

with $R_{M0} \in \mathbb{R}^{[M]\times V_M}$ given by $(R_{M0})_{[M]\times\langle M\rangle} = R_M$ and $(R_{M0})_{[M]\times(V_M\setminus\langle M\rangle)} = 0$ (the $[M] \times (V_M \setminus \langle M\rangle)$ zero matrix), is a well-defined bijection.

The inverse of this mapping takes the form

$$\begin{array}{ccc} \mathbf{P}(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M]) & \leftarrow & \mathbf{P}(\mathcal{U}) \\ (S_{V_M}, S_{[M\rangle\bullet}, S_{[M]\bullet}) & \leftarrow\!\shortmid & S \end{array} \tag{13}$$

with $S_{[M\rangle\bullet} := S_{[M\rangle} S_{\langle M\rangle}^{-1}$, $S_{[M]\bullet} := S_{[M]} - S_{[M\rangle} S_{\langle M\rangle}^{-1} S_{\langle M]}$, where $S_{\langle M\rangle}^{-1} := (S_{\langle M\rangle})^{-1}$.

Proof. The first claim follows from Proposition 3.1, Corollary 5.2, equation (8), and the fact

$$\begin{pmatrix} 1_{V_M} & 0 \\ -\Pi & 1_{[M]} \end{pmatrix}^{-1} = \begin{pmatrix} 1_{V_M} & 0 \\ \Pi & 1_{[M]} \end{pmatrix}$$

for all $\Pi \in \mathbb{R}^{[M]\times V_M}$. In order to establish (13) to be the inverse mapping to (12) we evaluate the mapping (12) at the point $(S_{V_M}, S_{[M\rangle\bullet}, S_{[M]\bullet})$ for a given matrix $S \in \mathbf{P}(\mathcal{U})$. The $V_M \times V_M$ submatrix of the image clearly is S_{V_M}. Due to $(\tilde{S}_{V_M})_{\langle M\rangle} = S_{\langle M\rangle}$ we have

$$S_{[M\rangle\bullet 0}\tilde{S}_{V_M} = \begin{pmatrix} S_{[M\rangle}S^{-1}_{\langle M\rangle} & 0 \end{pmatrix} \tilde{S}_{V_M} = \begin{pmatrix} S_{[M\rangle} & * \end{pmatrix},$$

and hence the $[M] \times V_M$ submatrix of the image is $\begin{pmatrix} S_{[M\rangle} & 0 \end{pmatrix} = S_{[M]\times V_M}$. Finally we have $S_{[M]\bullet} + S_{[M\rangle\bullet 0}\tilde{S}_{V_M} S^t_{[M\rangle\bullet 0} = S_{[M]\bullet} + S_{[M\rangle\bullet}S_{\langle M\rangle}S^t_{[M\rangle\bullet} = S_{[M]}$ which is why the $[M] \times [M]$ submatrix of the image is $S_{[M]}$. This completes the proof. $\square$

Following the notation from Proposition 5.2 we define $\Sigma^{-1}_{\langle B\rangle} := (\Sigma_{\langle B\rangle})^{-1}$, $\Sigma_{[B\rangle\bullet} := \Sigma_{[B\rangle}\Sigma^{-1}_{\langle B\rangle}$, $\Sigma_{[B]\bullet} := \Sigma_{[B]} - \Sigma_{[B\rangle}\Sigma^{-1}_{\langle B\rangle}\Sigma_{\langle B]}$, and $\Sigma^{-1}_{[B]\bullet} := (\Sigma_{[B]\bullet})^{-1}$ for all $\Sigma \in \mathbf{P}(\mathcal{U})$ and $B \in V/{\sim}$.

Corollary 5.3. *Let $S \in \mathbf{P}(\mathcal{U})$. Setting $D := \mathrm{Diag}(S_{[B]\bullet} \mid B \in V/{\sim})$ and $U_{[B]\times\langle B\rangle} := -S_{[B\rangle\bullet}$, $B \in V/{\sim}$, then yields the unique solution $(U, D) \in \mathbf{T}^1_\ell(\mathcal{V}) \times \mathbf{D}^+(\mathcal{V})$ to the equation $S = p\big(U^{-1}D(U^t)^{-1}\big)$, see Corollary 5.1.*

Proof. The proof is established by induction over $V/{\sim}$. Let $M \in V/{\sim}$ be a maximal box in $\mathcal{V}$ and assume S to be the image of $(S_M, R_M, L_M) \in \mathbf{P}(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M])$ under the mapping (12). Due to $\tilde{S}_M \in \mathbf{PD}(\mathcal{U}_M)$ we have the unique decomposition $\tilde{S}_M = U^{-1}_M D_M (U^t_M)^{-1}$ of $\tilde{S}_M$ with respect to $\mathcal{V}_M$, see Corollary 5.1. By the induction assumption we have $D_M = \mathrm{Diag}(S_{[B]\bullet} \mid B \in V_M/{\sim})$ and $(U_M)_{[B]\times\langle B\rangle} := -S_{[B\rangle\bullet}$, $B \in V_M/{\sim}$, since $(S_M)_{[B]\bullet}$ and $(S_M)_{[B\rangle\bullet}$ (taken with respect to $\mathcal{V}_M$) coincide with $S_{[B]\bullet}$ and $S_{[B\rangle\bullet}$ (taken with respect to $\mathcal{V}$), respectively, $B \in V_M/{\sim}$. Hence we have

$$\tilde{S} = \begin{pmatrix} (U_M)^{-1} & 0 \\ S_{[M\rangle\bullet 0} & 1_{[M]} \end{pmatrix} \begin{pmatrix} D_M & 0 \\ 0 & S_{[M]\bullet} \end{pmatrix} \begin{pmatrix} \big((U_M)^{-1}\big)^t & S^t_{[M\rangle\bullet 0} \\ 0 & 1_{[M]} \end{pmatrix}.$$

Due to

$$\begin{pmatrix} U_M & 0 \\ -S_{[M\rangle\bullet 0} & 1_{[M]} \end{pmatrix}^{-1} = \begin{pmatrix} U^{-1}_M & 0 \\ S_{[M\rangle\bullet 0} & 1_{[M]} \end{pmatrix} \in \mathbf{T}^1_\ell(\mathcal{V})$$

this proves the claim. $\square$

6. Calculation of a fundamental Jacobian.

We continue in the setting from Section 5.

Proposition 6.1. *Let $M \in V/{\sim}$ be a maximal box in $\mathcal{V}$. Then the Jacobian of the mapping (12) at $(S_M, R_M, L_M) \in \mathbf{P}(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M])$ and the Jacobian of its inverse mapping (13) at $S \in \mathbf{P}(\mathcal{U})$ are $|(S_M)_{\langle M\rangle}|^{[M]}$ and $|S_{\langle M\rangle}|^{-[M]}$, respectively.*

Proof. Let $S \in \mathbf{P}(\mathcal{U})$ be the image of $(S_M, R_M, L_M) \in \mathbf{P}(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M])$ under the mapping (12), that is,

$$S_{V_M} = S_M, \qquad S_{[M\rangle} = R_M(S_M)_{\langle M\rangle}, \qquad S_{[M]} = R_M S_{\langle M\rangle} R^t_M.$$

Then the Jacobian matrix of (12) at (S_M, R_M, L_M) evaluates to

$$\frac{\mathrm{d}(S_{V_M}, S_{[M\rangle}, S_{[M]})}{\mathrm{d}(S_M, R_M, L_M)} = \begin{array}{c} \\ S_{V_M} \\ S_{[M\rangle} \\ S_{[M]} \end{array} \begin{array}{c} \begin{array}{ccc} S_M & R_M & L_M \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ * & 1_{[M]} \otimes (S_M)_{\langle M\rangle} & 0 \\ * & * & 1 \end{pmatrix} \end{array},$$

where 1's represent identity matrices and asterisks represent blocks not further specified. The (absolute value of the) determinant of this matrix is $|(S_M)_{\langle M\rangle}|^{[M]}$, establishing the first claim. Expressing S_M, R_M, L_M in terms of S we have $(S_M)_{\langle M\rangle} = S_{\langle M\rangle}$. Hence the second claim follows. □

7. A fundamental integral.

As in the previous sections let $\mathcal{U}$ be a DUG with vertex set V and let $\mathcal{V}$ be a representation of $\mathcal{U}$ as an AMG. Let $\Delta \in \mathbf{PD}^0(\mathcal{U})$ and $\lambda \equiv (\lambda_B \,|\, B \in V/{\sim}) \in \mathbb{R}_+^{V/\sim}$. Then the integral

$$J_{\mathcal{V}}(\Delta, \lambda) := \int_{\mathbf{P}(\mathcal{U})} \prod \left(|S_{[B]\bullet}|^{\lambda_B} \,\middle|\, B \in V/{\sim} \right) \exp\{-\operatorname{tr}(\Delta S)\} \, \mathrm{d}\nu_{\mathcal{V}}(S),$$

with respect to the measure[1]

$$\mathrm{d}\nu_{\mathcal{V}}(S) := \prod \left(|S_{[B]\bullet}|^{-\frac{[B]+\langle B\rangle+1}{2}} \, |S_{\langle B\rangle}|^{-\frac{[B]}{2}} \,\middle|\, B \in V/{\sim} \right) \mathrm{d}S \tag{14}$$

is well-defined. We will derive a necessary and sufficient condition for the convergence of this integral and, given its convergence, evaluate the integral. To this end we use induction by $V/{\sim}$. Let $M \in V/{\sim}$ be a maximal box in $\mathcal{V}$. We use the mapping (13) to transform $J_{\mathcal{V}}(\Delta, \lambda)$ to an integral on the domain of the mapping (12). Note first that the measure $\nu_{\mathcal{V}}$ is transformed to the measure

$$|L_M|^{-\frac{[M]+\langle M\rangle+1}{2}} \, |(S_M)_{\langle M\rangle}|^{\frac{[M]}{2}} \, \mathrm{d}\big(\boldsymbol{\lambda}_{\mathbf{P}(\mathcal{U}_M)} \otimes \boldsymbol{\lambda}_{\mathbb{R}^{[M]\times\langle M\rangle}} \otimes \nu_{\mathcal{V}_M}\big)(S_M, R_M, L_M)$$

by Proposition 6.1, where $\boldsymbol{\lambda}_{\mathbf{P}(\mathcal{U}_M)}$ and $\boldsymbol{\lambda}_{\mathbb{R}^{[M]\times\langle M\rangle}}$ denote the standard Lebesgue measures on $\mathbf{P}(\mathcal{U}_M)$ and $\mathbb{R}^{[M]\times\langle M\rangle}$, respectively. Now we rewrite the function inside the integral $J_{\mathcal{V}}(\Delta, \lambda)$ in terms of the variables on the left-hand side of (12). We set

$$S = p\left(\begin{pmatrix} 1_{V_M} & 0 \\ R_{M0} & 1_M \end{pmatrix} \begin{pmatrix} \tilde{S}_M & 0 \\ 0 & L_M \end{pmatrix} \begin{pmatrix} 1_{V_M} & R_{M0}^t \\ 0 & 1_M \end{pmatrix} \right) \tag{15}$$

Furthermore we have the unique decomposition

$$\Delta = \begin{pmatrix} 1_{V_M} & -\Pi_{M0}^t \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} \Delta_M & 0 \\ 0 & \Upsilon_M \end{pmatrix} \begin{pmatrix} 1_{V_M} & 0 \\ -\Pi_{M0} & 1_{[M]} \end{pmatrix} \tag{16}$$

with $\Delta_M \in \mathbf{PD}^0(\mathcal{U}_M)$, $\Upsilon_M \in \mathbf{PD}([M])$, and $\Pi_{M0} \in \mathbb{R}^{[M]\times V_M}$ from Corollary 5.2. In particular we have $\Upsilon_M = \Delta_{[M]}$. Then we find

$$\begin{aligned}
\operatorname{tr}(\Delta S) &= \operatorname{tr}(\Delta \tilde{S}) \\
&= \operatorname{tr}\left(\begin{pmatrix} 1_{V_M} & -\Pi_{M0}^t \\ 0 & 1_{[M]} \end{pmatrix} \begin{pmatrix} \Delta_M & 0 \\ 0 & \Upsilon_M \end{pmatrix} \begin{pmatrix} 1_{V_M} & 0 \\ -\Pi_{M0} & 1_{[M]} \end{pmatrix} \right. \\
&\qquad \left. \times \begin{pmatrix} 1_{V_M} & 0 \\ R_{M0} & 1_{[M]} \end{pmatrix} \begin{pmatrix} \tilde{S}_M & 0 \\ 0 & L_M \end{pmatrix} \begin{pmatrix} 1_{V_M} & R_{M0}^t \\ 0 & 1_{[M]} \end{pmatrix} \right) \\
&= \operatorname{tr}\left(\begin{pmatrix} \Delta_M & 0 \\ 0 & \Upsilon_M \end{pmatrix} \begin{pmatrix} \tilde{S}_M & 0 \\ 0 & L_M + (R_{M0} - \Pi_{M0})\tilde{S}_M (R_{M0} - \Pi_{M0})^t \end{pmatrix} \right) \\
&= \operatorname{tr}(\Delta_M S_M) + \operatorname{tr}(\Upsilon_M L_M) + \operatorname{tr}\big(\Upsilon_M (R_M - \Pi_M)(S_M)_{\langle M\rangle}(R_M - \Pi_M)^t\big).
\end{aligned}$$

---

[1] This measure generalizes the invariant measure from Andersson & Wojnar (2004a) or Andersson & Wojnar (2004b) who consider the special case where $\mathbf{P}(\mathcal{U})$ is a homogeneous cone.

As a result the integral $J_{\mathcal{V}}(\Delta,\lambda)$ is transformed to

$$\int\limits_{\mathbf{P}(\mathcal{U}_M)}\int\limits_{\mathbb{R}^{[M]\times\langle M\rangle}}\int\limits_{\mathbf{PD}([M])} |L_M|^{\lambda_M-\frac{[M]+\langle M\rangle+1}{2}}\exp\{-\operatorname{tr}(\Upsilon_M L_M)\}$$
$$\times\,|(S_M)_{\langle M\rangle}|^{\frac{[M]}{2}}\exp\big\{-\operatorname{tr}\big(\Upsilon_M(R_M-\Pi_M)(S_M)_{\langle M\rangle}(R_M-\Pi_M^t)\big)\big\}$$
$$\times\prod\big(|(S_M)_{[B]\bullet}|^{\lambda_B}\,\big|\,B\in V_M/\sim\big)\exp\{-\operatorname{tr}(\Delta_M S_M)\}\,\mathrm{d}L_M\,\mathrm{d}R_M\,\mathrm{d}\nu_{\mathcal{V}_M}(S_M)$$
$$=\int\limits_{\mathbf{PD}([M])}|L_M|^{\lambda_M-\frac{[M]+\langle M\rangle+1}{2}}\exp\{-\operatorname{tr}(\Upsilon_M L_M)\}\,\mathrm{d}L_M$$
$$\times\int\limits_{\mathbf{P}(\mathcal{U}_M)}\int\limits_{\mathbb{R}^{[M]\times\langle M\rangle}}|(S_M)_{\langle M\rangle}|^{\frac{[M]}{2}}\exp\big\{-\operatorname{tr}\big(\Upsilon_M(R_M-\Pi_M)(S_M)_{\langle M\rangle}(R_M-\Pi_M^t)\big)\big\}$$
$$\times\prod\big(|(S_M)_{[B]\bullet}|^{\lambda_B}\,\big|\,B\in V_M/\sim\big)\exp\{-\operatorname{tr}(\Delta_M S_M)\}\,\mathrm{d}R_M\,\mathrm{d}\nu_{\mathcal{V}_M}(S_M).$$

The first of these two integrals converges if and only if $\lambda_M>\frac{[M]+\langle M\rangle-1}{2}$, and in that case its value is

$$\pi^{\frac{[M]\,([M]-1)}{4}}\prod\Big(\Gamma\big(\lambda_M-\tfrac{\langle M\rangle}{2}-\tfrac{i-1}{2}\big)\Big|\,i=1,\cdots,[M]\Big)\ |\Upsilon_M|^{-\lambda_M+\frac{\langle M\rangle}{2}}.$$

Integrating with respect to R_M the second integral above evaluates to

$$\pi^{\frac{[M]\langle M\rangle}{2}}\,|\Upsilon_M|^{-\frac{\langle M\rangle}{2}}\,J_{\mathcal{V}_M}(\Delta_M,\lambda_{-M}),$$

where $\lambda_{-M}:=(\lambda_B|\,B\in V_M/\sim)$. Due to $\Upsilon_M=\Delta_{[M]}=\Delta_{[M]^\circ}$ and $(\Delta_M)_{[B]^\circ}=\Delta_{[B]^\circ}$, $B\in V_M/\sim$, (see Remark 5.2) the induction assumption implies the integral $J_{\mathcal{V}}(\Delta,\lambda)$ to converge if and only if

$$\lambda_B>\tfrac{[B]+\langle B\rangle-1}{2},\qquad B\in V/\sim. \tag{17}$$

Furthermore, if this condition is satisfied, then we have

$$J_{\mathcal{V}}(\Delta,\lambda)=c_{\mathcal{V}}(\lambda)\prod\big(|\Delta_{[B]^\circ}|^{-\lambda_B}\,\big|\,B\in V/\sim\big),$$

where

$$c_{\mathcal{V}}(\lambda):=\pi^{\frac{\dim(\mathbf{P}(\mathcal{U}))-V}{2}}\prod\Big(\prod\Big(\Gamma\big(\lambda_B-\tfrac{\langle B\rangle}{2}-\tfrac{i-1}{2}\big)\Big|\,i=1,\cdots,[B]\Big)\Big|\,B\in V/\sim\Big).$$

8. The class of generalized Riesz distributions on $\mathbf{P}(\mathcal{U})$ associated with a decomposable undirected graph $\mathcal{U}$.

As in Section 7 let $\mathcal{U}$ be a DUG with vertex set V and let $\mathcal{V}$ be a representation of $\mathcal{U}$ as an AMG. Let $\lambda\equiv(\lambda_B|\,B\in V/\sim)\in\bigtimes\big(\big]\frac{[B]+\langle B\rangle-1}{2},\infty\big[\,\big|\,B\in V/\sim\big)$ as required in (17). From the previous section we obtain the full natural and canonical exponential family[1] on $\mathbf{P}(\mathcal{U})$ generated by the measure $\prod\big(|S_{[B]\bullet}|^{\lambda_B}\,\big|\,B\in V/\sim\big)\,\mathrm{d}\nu_{\mathcal{V}}(S)$, namely,

$$\big(\mathrm{R}_{\Delta,\lambda}\in\mathcal{P}(\mathbf{P}(\mathcal{U}))\big|\,\Delta\in\mathbf{PD}^0(\mathcal{U})\big)$$

[1]Formally this is an exponential family on the vector space $\mathbf{S}(\mathcal{U})$ concentrated on the open convex cone $\mathbf{P}(\mathcal{U})$.

with

$$\mathrm{dR}_{\Delta,\lambda}(S) := \frac{\pi^{\frac{V-\dim(\mathbf{P}(\mathcal{U}))}{2}} \prod\left(|\Delta_{B\circ}|^{\lambda_B}\,\middle|\, B\in V/\!\sim\right)\prod\left(|S_{B\bullet}|^{\lambda_B}\,\middle|\, B\in V/\!\sim\right)}{\prod\left(\prod\left(\Gamma\left(\lambda_B-\frac{\langle B\rangle}{2}-\frac{i-1}{2}\right)\middle|\, i=1,\cdots,[B]\right)\middle|\, B\in V/\!\sim\right)} \times \exp\{-\operatorname{tr}(\Delta S)\}\,\mathrm{d}\nu_{\mathcal{V}}(S).$$

Note that this exponential family is regular since $\mathbf{PD}^0(\mathcal{U})$ is an open subset of the vector space $\mathbf{S}(\mathcal{U})$. Also note that this family depends on the given representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG.

Definition 8.1. The probability measure $\mathrm{R}_{\Delta,\lambda}$ is called the *generalized Riesz distribution* on $\mathbf{P}(\mathcal{U})$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG, with *shape parameter* $\lambda \equiv (\lambda_B |\, B\in V/\!\sim)$ and *natural parameter* Δ.

The following proposition is an immediate consequence of the calculations in Section 7.

Proposition 8.1. *Let $M\in V/\!\sim$ be a maximal box and let $S\in\mathbf{P}(\mathcal{U})$ be a random element from $\mathrm{R}_{\Delta,\lambda}$, with $\Delta\in\mathbf{PD}^0(\mathcal{U})$ and $\lambda\equiv(\lambda_B|\, B\in V/\!\sim)$ satisfying* (17)*. Then we have:*

(i) The random elements $S_{[M]\bullet}\in\mathbf{PD}([M])$ and $(S_{[M\rangle\bullet}, S_{V_M})\in\mathbb{R}^{[M]\times\langle M\rangle}\times\mathbf{P}(\mathcal{U}_M)$ are independent.

(ii) The random element $S_{[M]\bullet}\in\mathbf{PD}([M])$ follows the classical Wishart distribution $\mathrm{W}_{\Delta_{[M]},\lambda_M-\frac{\langle M\rangle}{2}}$ with shape parameter $\lambda_M-\frac{\langle M\rangle}{2}$ and natural parameter $\Delta_{[M]}$.

(iii) The distribution of the random element $(S_{[M\rangle\bullet}, S_{V_M})\in\mathbb{R}^{[M]\times\langle M\rangle}\times\mathbf{P}(\mathcal{U}_M)$ is described as follows: The conditional distribution of $S_{[M\rangle\bullet}$ given S_{V_M} is $\mathrm{N}_{\Pi_M,(2\Delta_{[M]}\otimes S_{\langle M\rangle})^{-1}}$, the normal distribution on $\mathbb{R}^{[M]\times\langle M\rangle}$ with expectation Π_M (see equation (16)*) and precision $2\Delta_{[M]}\otimes S_{\langle M\rangle}$; in particular, this conditional distribution depends on S_{V_M} only through $S_{\langle M\rangle}$. The distribution of S_{V_M} is the generalized Riesz distribution $\mathrm{R}_{\Delta_M,\lambda_{-M}}$ on $\mathbf{P}(\mathcal{U}_M)$ with respect to the representation $\mathcal{V}_M$ of $\mathcal{U}_M$ as an AMG, with natural parameter Δ_M (see equation* (16)*) and shape parameter $\lambda_{-M}:=(\lambda_B|\, B\in V_M/\!\sim)$.*

9. The generalization of the λ-matrix inverse mapping.

Again let $\mathcal{U}$ be a DUG with vertex set V and let $\mathcal{V}$ be a representation of $\mathcal{U}$ as an AMG. Let $\Sigma\in\mathbf{P}(\mathcal{U})$ and $\lambda\equiv(\lambda_B|\, B\in V/\!\sim)\in\mathbb{R}_+^{V/\sim}$. Writing $\Sigma = p\big(U^{-1}\operatorname{Diag}(\Sigma_{[B]\bullet}|\, B\in V/\!\sim)(U^t)^{-1}\big)$ with the uniquely determined matrix $U\in\mathbf{T}^1_\ell(\mathcal{V})$ (see Corollary 5.3) we may define $\Sigma^{-\lambda} := U^t\operatorname{Diag}(\lambda_B\Sigma^{-1}_{[B]\bullet}|\, B\in V/\!\sim)\,U\in\mathbf{PD}^0(\mathcal{U})$, the λ-*inverse* of Σ. The mapping

$$\begin{array}{ccc}\mathbf{P}(\mathcal{U}) & \to & \mathbf{PD}^0(\mathcal{U}),\\ \Sigma & \mapsto & \Sigma^{-\lambda},\end{array}$$

clearly is a bijection with inverse mapping

$$\begin{array}{rcl}\mathbf{P}(\mathcal{U}) & \leftarrow & \mathbf{PD}^0(\mathcal{U}),\\ p\big(U^{-1}\operatorname{Diag}(\lambda_B\Delta^{-1}_{[B]\circ}|\, B\in V/\!\sim)(U^t)^{-1}\big) & \leftarrow\!\shortmid & \Delta = U^t\operatorname{Diag}(\Delta_{[B]\circ}|\, B\in V/\!\sim)\,U.\end{array}$$

Note

$$(\Sigma^{-\lambda})_{[B]\circ} = \lambda_B\Sigma^{-1}_{[B]\bullet},\quad B\in V/\!\sim. \tag{18}$$

Also note that $\Sigma^{-\lambda}$ depends on the given representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG although our notation does not reflect that dependence. As in Section 2 we define $\Sigma^{-\mu} := \Sigma^{-(\mu|\, B\in V/\sim)}$ for all $\Sigma \in \mathbf{P}(\mathcal{U})$, $\mu \in \mathbb{R}_+$. In the case $\mu = 1$ this notation is consistent with (8).

Remark 9.1. If $\Delta = \Sigma^{-\lambda}$, then $E := \mathrm{Diag}(\lambda_B \Sigma^{-1}_{[B]\bullet} |\, B \in V/\sim)$ and $U_{[B]\times\langle B\rangle} := -\Sigma_{[B\rangle\bullet}$, $B \in V/\sim$, yield the unique solution $(U, E) \in \mathbf{T}^1_\ell(\mathcal{V}) \times \mathbf{D}^+(\mathcal{V})$ to the equation $\Delta = U^t EU$, see Corollaries 5.1 and 5.3.

10. The expectation of a generalized Riesz distribution.

We continue in the setting from the previous sections.

Proposition 10.1. *Let $\Delta \in \mathbf{PD}^0(\mathcal{U})$, let $\lambda \equiv (\lambda_B |\, B \in V/\sim)$ satisfy* (17)*, and let $\Sigma \in \mathbf{P}(\mathcal{U})$ satisfy $\Delta = \Sigma^{-\lambda}$. Then we have $\mathbb{E}\,(\mathrm{R}_{\Delta,\lambda}) = \Sigma$.*

Proof. Clearly we have $\mathbb{E}\,(\mathrm{R}_{\Delta,\lambda}) = \int_{\mathbf{P}(\mathcal{U})} S\,\mathrm{dR}_{\Delta,\lambda}(S) = p\big(\int_{\mathbf{P}(\mathcal{U})} \tilde S\,\mathrm{dR}_{\Delta,\lambda}(S)\big)$. We will use induction on $V/\sim$ to evaluate the integral $\int_{\mathbf{P}(\mathcal{U})} \tilde S\,\mathrm{dR}_{\Delta,\lambda}(S)$. Let $M \in V/\sim$ be a maximal box in $\mathcal{V}$. Using the mapping (13) we may transform the given integral on $\mathbf{P}(\mathcal{U})$ into one on $\mathbf{P}(\mathcal{U}_M) \times \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{PD}([M])$. Furthermore we utilize the representations of S and Δ given in (15) and (16). Arguments similar to those in the proof of Corollary 5.3 (see also Remark 9.1) show the assumption $\Delta = \Sigma^{-\lambda}$ to imply $\Delta_M = (\Sigma_{V_M})^{-\lambda_{-M}}$, $\Pi_M = \Sigma_{[M\rangle\bullet}$, and $\Upsilon_M = \Delta_{[M]} = \lambda_M \Sigma^{-1}_{[M]\bullet}$. Then the $V_M \times V_M$ submatrix of $\int_{\mathbf{P}(\mathcal{U})} \tilde S\,\mathrm{dR}_{\Delta,\lambda}(S)$ turns into

$$\iiint \tilde S_M \,\mathrm{dW}_{\Delta_{[M]},\lambda_M - \frac{\langle M\rangle}{2}}(L_M)\,\mathrm{dN}_{\Pi_M,\left(2\Delta_{[M]}\otimes (S_M)_{\langle M\rangle}\right)^{-1}}(R_M)\,\mathrm{dR}_{\Delta_M,\lambda_{-M}}(S_M)$$
$$= \int\limits_{\mathbf{P}(\mathcal{U}_M)} \tilde S_M\,\mathrm{dR}_{\Delta_M,\lambda_{-M}}(S_M) = \tilde\Sigma_{V_M},$$

where the last equality is due to the induction assumption. By the same argument the $[M] \times V_M$ submatrix of $\int_{\mathbf{P}(\mathcal{U})} \tilde S\,\mathrm{dR}_{\Delta,\lambda}(S)$ is

$$\iiint R_{M0}\tilde S_M \,\mathrm{dW}_{\Delta_{[M]},\lambda_M - \frac{\langle M\rangle}{2}}(L_M)\,\mathrm{dN}_{\Pi_M,\left(2\Delta_{[M]}\otimes (S_M)_{\langle M\rangle}\right)^{-1}}(R_M)\,\mathrm{dR}_{\Delta_M,\lambda_{-M}}(S_M)$$
$$= \Pi_{M0} \int\limits_{\mathbf{P}(\mathcal{U}_M)} \tilde S_M\,\mathrm{dR}_{\Delta_M,\lambda_{-M}}(S_M) = \Sigma_{[M\rangle\bullet}\tilde\Sigma_{V_M} = \tilde\Sigma_{[M]\times V_M},$$

again using the induction assumption. Finally, the $[M] \times [M]$ submatrix of the integral $\int_{\mathbf{P}(\mathcal{U})} \tilde S\,\mathrm{dR}_{\Delta,\lambda}(S)$ is

$$\int\limits_{\mathbf{PD}([M])} L_M\,\mathrm{dW}_{\Delta_{[M]},\lambda_M - \frac{\langle M\rangle}{2}}(L_M)$$
$$+ \int\limits_{\mathbf{P}(\mathcal{U}_M)} \int\limits_{\mathbb{R}^{[M]\times\langle M\rangle}} R_M (S_M)_{\langle M\rangle} R_M^t \,\mathrm{dN}_{\Pi_M,\left(2\Delta_{[M]}\otimes (S_M)_{\langle M\rangle}\right)^{-1}}(R_M)\,\mathrm{dR}_{\Delta_M,\lambda_{-M}}(S_M)$$
$$= \left(\lambda_M - \tfrac{\langle M\rangle}{2}\right)\Delta^{-1}_{[M]} + \int\limits_{\mathbf{P}(\mathcal{U}_M)} \left(\tfrac{\langle M\rangle}{2}\Delta^{-1}_{[M]} + \Pi_M (S_M)_{\langle M\rangle}\Pi_M^t\right)\mathrm{dR}_{\Delta_M,\lambda_{-M}}(S_M)$$

$$= \lambda_M \Delta_{[M]}^{-1} + \Pi_M \int\limits_{\mathbf{P}(\mathcal{U}_M)} (S_M)_{\langle M\rangle} \, \mathrm{dR}_{\Delta_M, \lambda_{-M}}(S_M) \, \Pi_M^t$$

$$= \Sigma_{[M]\bullet} + \Sigma_{[M\rangle\bullet}(\Sigma_{V_M})_{\langle M\rangle}\Sigma_{[M\rangle\bullet}^t = \Sigma_{[M]} = \tilde{\Sigma}_{[M]},$$

using the induction assumption once more. This completes the proof. □

Replacing $\Delta \in \mathbf{PD}^0(\mathcal{U})$ by $\Sigma^{-\lambda}$ with $\Sigma \in \mathbf{P}(\mathcal{U})$ we obtain the generalized Riesz distribution $\mathbb{R}_{\Sigma,\lambda}$ parameterized by its expectation. Due to $(\Sigma^{-\lambda})_{[B]\circ} = \lambda_B(\Sigma_{[B]\bullet})^{-1}$, $B \in V/\sim$, we may represent this distribution as

$$\mathrm{d}\mathbb{R}_{\Sigma,\lambda}(S) := \frac{\pi^{\frac{\dim(\mathbf{P}(\mathcal{U}))-V}{2}} \prod \left(\lambda_B^{\lambda_B[B]} \middle|\, B \in V/\sim\right)}{\prod \Big(\prod \left(\Gamma(\lambda_B - \frac{\langle B\rangle}{2} - \frac{i-1}{2}) \middle|\, i = 1, \cdots, [B]\right) \Big|\, B \in V/\sim \Big)}$$
$$\times \frac{\prod \left(|S_{B\bullet}|^{\lambda_B} \,\middle|\, B \in V/\sim\right)}{\prod \left(|\Sigma_{B\bullet}|^{\lambda_B} \,\middle|\, B \in V/\sim\right)} \exp\{-\operatorname{tr}(\Sigma^{-\lambda}S)\} \, \mathrm{d}\nu_{\mathcal{V}}(S).$$

Definition 10.1. The probability measure $\mathbb{R}_{\Sigma,\lambda}$ is called the *generalized Riesz distribution* on $\mathbf{P}(\mathcal{U})$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG with *shape parameter* $\lambda \equiv (\lambda_B |\, B \in V/\sim)$ and *expectation parameter* Σ.

The *Riesz model given by* $\mathcal{V}$ (with fixed shape parameter λ) in its expectation parameterization is then $\left(\mathbb{R}_{\Sigma,\lambda} \in \mathcal{P}(\mathbf{P}(\mathcal{U})) \middle|\, \Sigma \in \mathbf{P}(\mathcal{U})\right)$. It is trivial that the ML estimator $\tilde{\Sigma}(S)$ for $\Sigma \in \mathbf{P}(\mathcal{U})$ at the observation point $S \in \mathbf{P}(\mathcal{U})$ exists for all $S \in \mathbf{P}(\mathcal{U})$ and is uniquely given by $\hat{\Sigma}(S) = S$. It is also trivial that $\hat{\Sigma}$ is complete and sufficient.

In the expectation parameterization Proposition 8.1 takes the following form.

Proposition 10.2. *Let $S \in \mathbf{P}(\mathcal{U})$ be a random element from $\mathbb{R}_{\Sigma,\lambda}$, with $\Sigma \in \mathbf{P}(\mathcal{U})$ and $\lambda \equiv (\lambda_B |\, B \in V/\sim)$ satisfying (17). Let $M \in V/\sim$ be a maximal box. Then we have:*

(i) The random elements $S_{[M]\bullet} \in \mathbf{PD}([M])$ and $(S_{[M\rangle\bullet}, S_{V_M}) \in \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{P}(\mathcal{U}_M)$ are independent.

(ii) The random element $S_{[M]\bullet} \in \mathbf{PD}([M])$ has the classical Wishart distribution $\mathbb{W}_{\Sigma_{[M]\bullet}, \lambda_M - \frac{\langle M\rangle}{2}}$ with shape parameter $\lambda_M - \frac{\langle M\rangle}{2}$ and expectation parameter $\Sigma_{[M]\bullet}$.

(iii) The distribution of the random element $(S_{[M\rangle\bullet}, S_{V_M}) \in \mathbb{R}^{[M]\times\langle M\rangle} \times \mathbf{P}(\mathcal{U}_M)$ is described as follows: The conditional distribution of $S_{[M\rangle\bullet}$ given S_{V_M} is $\mathrm{N}_{\Sigma_{[M\rangle\bullet}, \frac{1}{2\lambda_M}\Sigma_{[M]\bullet}\otimes S^{-1}_{\langle M\rangle}}$, the normal distribution on $\mathbb{R}^{[M]\times\langle M\rangle}$ with expectation $\Sigma_{[M\rangle\bullet}$ and variance matrix $\frac{1}{2\lambda_M}\Sigma_{[M]\bullet} \otimes S^{-1}_{\langle M\rangle}$; in particular this conditional distribution depends on S_{V_M} only through $S_{\langle M\rangle}$. The distribution of S_{V_M} is $\mathbb{R}_{\Sigma_{V_M},\lambda_{-M}}$, the generalized Riesz distribution on $\mathbf{P}(\mathcal{U}_M)$ with expectation parameter Σ_{V_M}, (see equation (13)) and shape parameter λ_{-M}.

11. Marginalization, decomposition, product, and convolution of Riesz distributions.

We continue in the setting from sections 8 and 10. Let $A \subseteq V/\!\sim$ be an ancestral subset[1] and set $V_A := \dot{\bigcup}\big([B]\big|\, B \in A\big)$. Let $\mathcal{V}_A$ and $\mathcal{U}_A$ be the AMG and DUG induced by the subset $V_A \subseteq V$ in $\mathcal{V}$ and $\mathcal{U}$, respectively. Then $\mathcal{V}_A$ is a representation of the DUG $\mathcal{U}_A$. The mapping

$$\begin{array}{rcl} \mathbf{P}(\mathcal{U}) & \to & \mathbf{P}(\mathcal{U}_A), \\ S & \mapsto & S_{V_A}, \end{array}$$

is then well-defined and onto.

Proposition 11.1. *(Marginalization) Let the random element $S \in \mathbf{P}(\mathcal{U})$ follow the generalized Riesz distribution $\mathbb{R}_{\Sigma,\lambda}$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG, with expectation parameter $\Sigma \in \mathbf{P}(\mathcal{U})$ and shape parameter $\lambda \equiv (\lambda_B|\, B \in V/\!\sim)$. Then the random element $S_{V_A} \in \mathbf{P}(\mathcal{U}_A)$ follows the generalized Riesz distribution $\mathbb{R}_{\Sigma_{V_A},\lambda_A}$, where $\lambda_A := (\lambda_B|\, B \in A)$.*

Proof. The claim is readily verified by applying the last part of Proposition 10.2 (iii) several times. □

Our decomposition, product, and convolution results (of which the first two are trivial) will be formulated for two components only. The results extend trivially to finitely many items. Let $\mathcal{U} = \mathcal{U}_1 \,\dot\cup\, \mathcal{U}_2$ be a decomposition of $\mathcal{U} = (V, F)$ into two components, that is, $\mathcal{U}_1 = (V_1, F_1)$ and $\mathcal{U}_2 = (V_2, F_2)$ with $V = V_1 \,\dot\cup\, V_2$ and $F = F_1 \dot\cup F_2$. Since $\mathcal{V}$ is a representation of $\mathcal{U}$ as an AMG the AMG $\mathcal{V}_i$ induced in $\mathcal{V}$ by V_i is a representations of $\mathcal{U}_i$ as an AMG, $i = 1, 2$. Note the identities $\mathcal{V} = \mathcal{V}_1 \,\dot\cup\, \mathcal{V}_2$ and $V/\!\sim\, = (V_1/\!\sim)\dot\cup(V_2/\!\sim)$. Due to $\mathbf{P}(\mathcal{U}) = \big\{\operatorname{Diag}(\Sigma_1, \Sigma_2)\big|\,(\Sigma_1, \Sigma_2) \in \mathbf{P}(\mathcal{U}_1) \times \mathbf{P}(\mathcal{U}_2)\big\}$ we also have the identity of convex cones $\mathbf{P}(\mathcal{U}_1 \,\dot\cup\, \mathcal{U}_2) = \mathbf{P}(\mathcal{U}_1) \times \mathbf{P}(\mathcal{U}_2)$. The following proposition is an immediate consequence of these facts.

Proposition 11.2. *(Decomposition and product) Let $\mathcal{U} = \mathcal{U}_1 \,\dot\cup\, \mathcal{U}_2$ be a decomposition of $\mathcal{U}$ as above. Let $\mathbb{R}_{\Sigma,\lambda}$ be the generalized Riesz distribution on $\mathbf{P}(\mathcal{U})$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG, with expectation parameter $\Sigma \in \mathbf{P}(\mathcal{U})$ and shape parameter $\lambda \equiv (\lambda_B|\, B \in V/\!\sim)$. Then we have*

$$\mathbb{R}_{\Sigma,\lambda} = \mathbb{R}_{\Sigma_1,\lambda_1} \otimes \mathbb{R}_{\Sigma_2,\lambda_2}, \tag{19}$$

where $\Sigma_1 := \Sigma_{V_1} \in \mathbf{P}(\mathcal{U}_1)$, $\lambda_1 := (\lambda_B|\, B \in V_1/\!\sim)$, $\Sigma_2 := \Sigma_{V_2} \in \mathbf{P}(\mathcal{U}_2)$, and $\lambda_2 := (\lambda_B|\, B \in V_2/\!\sim)$.

Conversely, if $\mathbb{R}_{\Sigma_i,\lambda_i}$ is the generalized Riesz distribution on $\mathbf{P}(\mathcal{U}_i)$ with respect to the representation $\mathcal{V}_i$ of $\mathcal{U}_i$ as an AMG, with expectation parameter $\Sigma_i \in \mathbf{P}(\mathcal{U}_i)$ and shape parameter $\lambda_i \equiv (\lambda_B|\, B \in V_i/\!\sim)$, $i = 1, 2$, then equation (19) holds with $\Sigma := \operatorname{Diag}(\Sigma_1, \Sigma_2)$ and $\lambda := (\lambda_B|\, B \in V/\!\sim)$.

Proposition 11.3. *(Convolution) Let $\mathrm{R}_{\Delta,\lambda_i}$ be the generalized Riesz distribution on $\mathbf{P}(\mathcal{U})$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG, with natural parameter $\Delta \in \mathbf{PD}^0(\mathcal{U})$ and shape parameter $\lambda_i \equiv (\lambda_{iB}|\, B \in V/\!\sim)$, $i = 1, 2$. Then we have*

$$\mathrm{R}_{\Delta,\lambda_1} * \mathrm{R}_{\Delta,\lambda_2} = \mathrm{R}_{\Delta,\lambda_1+\lambda_2}.$$

---

[1] A subset $A \subseteq V/\!\sim$ is *ancestral* in the ADG of boxes $\mathcal{V}/\!\sim$ if $B \in A$ and $B' \in V/\!\sim$ with $B' \to B$ implies $B' \in A$.

In other words: If $S_1, S_2 \in \mathbf{P}(\mathcal{U})$ are two independent random elements, S_1 following the generalized Riesz distribution $\mathrm{R}_{\Delta,\lambda_1}$ and S_2 following the generalized Riesz distribution $\mathrm{R}_{\Delta,\lambda_2}$ (both generalized Riesz distributions with the same natural parameter Δ), then S_1+S_2 follows the generalized Riesz distribution $\mathrm{R}_{\Delta,\lambda_1+\lambda_2}$ with respect to the representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG, with natural parameter $\Delta \in \mathbf{PD}^0(\mathcal{U})$ and shape parameter $\lambda_1 + \lambda_2$.

Proof. The multivariate Laplace transform for $\mathrm{R}_{\Delta,\lambda_i}$ at the point $T \in \mathbf{S}(\mathcal{U})$ with $\Delta - T \in \mathbf{PD}^0(\mathcal{U})$ is given by

$$\frac{\prod\big(|\Delta_{B\circ}|^{\lambda_{iB}}\,\big|\,B \in V/{\sim}\big)}{\prod\big(|(\Delta-T)_{B\circ}|^{\lambda_{iB}}\,\big|\,B \in V/{\sim}\big)}\,, \qquad i = 1, 2.$$

The product of these two Laplace transforms is

$$\frac{\prod\big(|\Delta_{B\circ}|^{\lambda_{1B}+\lambda_{2B}}\,\big|\,B \in V/{\sim}\big)}{\prod\big(|(\Delta-T)_{B\circ}|^{\lambda_{1B}+\lambda_{2B}}\,\big|\,B \in V/{\sim}\big)}\,,$$

the Laplace transform of $\mathrm{R}_{\Delta,\lambda_1+\lambda_2}$ at the point $T \in \mathbf{S}(\mathcal{U})$ with $\Delta - T \in \mathbf{PD}^0(\mathcal{U})$. Since $\big\{T \in \mathbf{S}(\mathcal{U})\,\big|\,\Delta - T \in \mathbf{PD}^0(\mathcal{U})\big\}$ is open in $\mathbf{S}(\mathcal{U})$ the claim follows. $\square$

12. The connection to the work by Letac and Massam (2007)

In this section we investigate the relation of our work to the work by Letac & Massam (2007), abbreviated here as LM. Let $\mathcal{U}$ be a decomposable undirected graph (DUG). In the paper by LM, titled *Wishart distributions for decomposable graphs*, the authors define *Wishart distributions of type I* on the open convex cone $Q_{\mathcal{U}} \equiv \mathbf{P}(\mathcal{U})$, connecting LM's notation $Q_{\mathcal{U}}$ to our notation $\mathbf{P}(\mathcal{U})$. Now we recall, in fact almost quote, LM's definitions and results surrounding their Wishart distributions of type I.

Let $\mathcal{C}$, $\mathcal{S}$, and $\nu(S)$ denote the set of cliques in $\mathcal{U}$, the set of separators in $\mathcal{U}$, and the multiplicity of a separator $S \in \mathcal{S}$, respectively[1]. As in LM we shall now assume $\mathcal{C} > 1$. The central idea, stated in equation (3.1) in LM and reproduced here in LM's notation, is the study of the integral

$$I(\alpha, \beta, \Delta) := \int_{\mathbf{P}(\mathcal{U})} H(\alpha, \beta, X)\, \exp\{-\operatorname{tr}(\Delta X)\}\, \mathrm{d}\mu_{\mathcal{U}}(X),$$

with $(\alpha, \beta) \equiv \big((\alpha_C\,|\,C \in \mathcal{C}), (\beta_S\,|\,S \in \mathcal{S})\big) \in \mathbb{R}^{\mathcal{C}} \times \mathbb{R}^{\mathcal{S}}$, $\Delta \in \mathcal{P}_{\mathcal{U}} \equiv \mathbf{PD}^0(\mathcal{U})$, where we have renamed y to Δ and x to X, and where

$$H(\alpha, \beta, X) := \frac{\prod\big(|X_C|^{\alpha_C}\,\big|\,C \in \mathcal{C}\big)}{\prod\big(|X_S|^{\nu(S)\beta_S}\,\big|\,S \in \mathcal{S}\big)}\,, \quad X \in \mathbf{P}(\mathcal{U}),$$

and

$$\mathrm{d}\mu_{\mathcal{U}}(X) := H\Big(\big(-\tfrac{C+1}{2}\,\big|\,C \in \mathcal{C}\big),\ \big(-\tfrac{S+1}{2}\,\big|\,S \in \mathcal{S}\big),\ X\Big)\,\mathrm{d}X. \tag{20}$$

The set of all $(\alpha, \beta) \in \mathbb{R}^{\mathcal{C}} \times \mathbb{R}^{\mathcal{S}}$ with $I(\alpha, \beta, \Delta) < \infty$ and $I(\alpha, \beta, \Delta)/H\big(\alpha, \beta, p(\Delta^{-1})\big) = c_I(\alpha, \beta)$, $\Delta \in \mathbf{PD}^0(\mathcal{U})$, (where the right-hand side does not depend on Δ) is

[1]See Lauritzen (1996, Chapter 2) for these concepts from graph theory.

denoted by $\mathcal{A}$ in LM. Then LM define[1] the *Wishart distribution of type I* with parameters $\big((\alpha,\beta),\Delta\big) \in \mathcal{A} \times \mathbf{PD}^0(\mathcal{U})$ as

$$\mathrm{d}W_{\mathbf{P}(\mathcal{U}),\alpha,\beta,\Delta}(X) := \frac{1}{c_I(\alpha,\beta)\, H\big(\alpha,\beta,p(\Delta^{-1})\big)}\, H(\alpha,\beta,X)\, \exp\{-\operatorname{tr}(\Delta X)\}\, \mathrm{d}\mu_{\mathcal{U}}(X),$$

The problem of characterizing $\mathcal{A}$ and calculating $c_I(\alpha,\beta)$, $(\alpha,\beta) \in \mathcal{A}$ is a main consideration in LM. Nevertheless LM do not obtain a complete solution to this problem.

If $\mathcal{U}$ does not contain the DUG $\bullet - \bullet - \bullet - \bullet$ (denoted by A_4 in LM) as an induced subgraph, then the open convex cone $\mathbf{P}(\mathcal{U})$ is a homogeneous cone. In this case LM's family of Wishart distributions of type I is identical[2] to the family of *general Wishart distributions* on $\mathbf{P}(\mathcal{U})$ obtained by Andersson & Wojnar (2004a) and Andersson & Wojnar (2004b) for any homogeneous cone, and hence for the homogeneous cone $\mathbf{P}(\mathcal{U})$. LM's main problem is thus already solved in this special case. Nevertheless, LM presents a self-contained alternative version of the solution in this special case, cf. Sections 2.2 and 3.3 in LM.

In the non-homogeneous case, that is, if $\mathcal{U}$ does contain A_4 as an induced subgraph, LM's solution to the above problem is incomplete, as they also point out themselves. In that case LM define subsets $\mathcal{A}_P \subseteq \mathcal{A}$ (see below), where each subset A_P depends on a perfect ordering $P \equiv (C_1,\cdots,C_{\mathcal{C}})$ of the cliques of $\mathcal{U}$; however, the mapping $P \mapsto \mathcal{A}_P$ is not injective in general. Each of these subsets is a convex cone of dimension $\mathcal{C}+1$. Clearly we have $\bigcup_P \mathcal{A}_P \subseteq \mathcal{A}$. Since it is unknown whether equality holds or not the set $\mathcal{A}$ is not characterized by LM[3]. Next LM consider the family of Wishart distributions of type I parameterized by $\mathcal{A}_P \times \mathbf{PD}^0(\mathcal{U})$, for all perfect orderings P as above. These families do in fact depend on the respective perfect orderings, and each such family is closed under convolution. Replacing the individual sets $\mathcal{A}_P$ with $\bigcup_P \mathcal{A}_P$ LM obtain a family of Wishart distributions of type I that does not depend on any perfect ordering of the cliques of $\mathcal{U}$ but fails to be closed under convolution.

Let $P \equiv (C_1,\cdots,C_{\mathcal{C}})$ be a perfect ordering of the cliques of $\mathcal{U}$ as above and let $S_2,S_3,\cdots,S_{\mathcal{C}}$ be the ordered listing of separators (with possible repetitions) induced by P. Then LM define $\mathcal{A}_P$ as the set of all $(\alpha,\beta) \equiv \big((\alpha_C|\, C\in\mathcal{C}),(\beta_S|\, S\in\mathcal{S})\big) \in \mathbb{R}^{\mathcal{C}} \times \mathbb{R}^{\mathcal{S}}$ such that

$$\begin{gathered}\alpha_C > \tfrac{C-1}{2}\,, \quad C\in\mathcal{C},\\ \alpha_{C_1}+\delta_2 > \tfrac{S_2-1}{2}\,,\\ \sum\big(\alpha_{C_q}\big|\, q\in J_P(S)\big) - \nu(S)\,\beta_S = 0, \qquad S\in\mathcal{S},\ S\neq S_2,\end{gathered} \tag{21}$$

where $J_P(S) = \big\{j\in\{1,\cdots,\mathcal{C}\}\big|\, S_j = S\big\}$ and $\delta_2 := \sum\big(\alpha_{C_q}\big|\, q\in J_P(S_2)\big) - \nu(S_2)\beta_2$, see Section 3.4 in LM. Furthermore LM obtain, for all $(\alpha,\beta)\in\mathcal{A}_P$,

$$c_I(\alpha,\beta) = \Gamma_{S_2}(\alpha_1+\delta_2)\,\frac{\Gamma_{C_1}(\alpha_1)}{\Gamma_{S_2}(\alpha_1)}\prod\left(\frac{\Gamma_{C_q}(\alpha_q)}{\Gamma_{S_q}(\alpha_q)}\middle|\, q=2,\cdots,\mathcal{C}\right)$$

[1]LM also state the Wishart distribution of type I using the parameter $\Sigma := p(\Delta^{-1}) \in \mathbf{P}(\mathcal{U})$. This parameter is different from our expectation parameter (also denoted by Σ).

[2]Identity holds up to a trivial reparameterization.

[3]Nevertheless LM do establish equality in the case $\mathcal{U} = A_4$. Furthermore it is known that equality does not hold in general for the homogeneous case.

with

$$\Gamma_r(p) := \pi^{\frac{r\,(r-1)}{4}} \prod \left(\Gamma\big(p - \tfrac{j-1}{2}\big) \,\big|\, j = 1, \cdots, r\right),$$

$p \in \mathbb{R}_+$, $r \in \mathbb{N} := \{1, 2, \cdots\}$, $p > \frac{r-1}{2}$.

We shall now establish LM's family of Wishart distributions of type I parameterized by $\mathcal{A}_P \times \mathbf{PD}^0(\mathcal{U})$ to be a family of generalized Riesz distributions (see Definition 8.1) with respect to one specific representation of $\mathcal{U}$ as an AMG. Since P is a perfect ordering we have the *histories* $H_j := C_1 \cup C_2 \cup \cdots \cup C_j$, $j = 1, \cdots, \mathcal{C}$, the *separators* $S_j := C_j \cap H_{j-1}$, $j = 2, \cdots, \mathcal{C}$, and the *remainders* $R_j = C_j \setminus H_{j-1}$, $j = 2, \cdots, \mathcal{C}$. Now we replace all undirected edges between vertices $v \in S_j$ and $v' \in R_j$ with arrows $v \to v'$. We also replace all undirected edges between vertices $v \in S_2$ and $v' \in C_1 \setminus S_2$ by the arrow $v \to v'$. Using the definition of perfect orderings this assignment is readily seen to be consistent and to result in an AMG $\mathcal{V}_P$ without triplexes and with $\mathcal{C} + 1$ boxes $S_2, C_1 \setminus S_2, R_2, \cdots, R_{\mathcal{C}}$, all of which are complete. As a consequence $\mathcal{V}_P$ is a representation of $\mathcal{U}$ as an AMG, see Definition 4.1. Furthermore we have $\langle S_2 \rangle = \varnothing$, $\langle C_1 \setminus S_2 \rangle = S_2$, and $\langle R_j \rangle = S_j$, $j = 2, \cdots, \mathcal{C}$. Abbreviating $\mathcal{V} := \mathcal{V}_P$ and replacing S with X our measure $\nu_{\mathcal{V}}$ from equation (14) is then

$$\begin{aligned}
&\mathrm{d}\nu_{\mathcal{V}}(X) \\
&:= \prod \left(|X_{[B]\bullet}|^{-\frac{[B]+\langle B\rangle+1}{2}} \, |X_{\langle B\rangle}|^{-\frac{[B]}{2}} \,\Big|\, B \in V/{\sim} \right) \mathrm{d}X \\
&= |X_{S_2}|^{-\frac{S_2+1}{2}} \, |X_{(C_1\setminus S_2)\bullet}|^{-\frac{C_1\setminus S_2+S_2+1}{2}} \, |X_{S_2}|^{-\frac{C_1\setminus S_2}{2}} \, |X_{S_2}|^{-\frac{C_1+1}{2}} \, |X_{(C_1\setminus S_2)\bullet}|^{-\frac{C_1+1}{2}} \\
&\quad \times \prod \left(|X_{R_j\bullet}|^{-\frac{R_j+S_j+1}{2}} \, |X_{S_j}|^{-\frac{R_j}{2}} \,\Big|\, j = 2, \cdots, \mathcal{C} \right) \mathrm{d}X \\
&= |X_{C_1}|^{-\frac{C_1+1}{2}} \prod \left(\big(|X_{R_j\bullet}| \, |X_{S_j}| \big)^{-\frac{R_j+S_j+1}{2}} \, |X_{S_j}|^{\frac{S_j+1}{2}} \,\Bigg|\, j = 2, \cdots, \mathcal{C} \right) \mathrm{d}X \\
&= \frac{\prod \left(|X_{C_j}|^{-\frac{C_j+1}{2}} \,\Big|\, j = 1, \cdots, \mathcal{C} \right)}{\prod \left(|X_{S_j}|^{-\frac{S_j+1}{2}} \,\Big|\, j = 2, \cdots, \mathcal{C} \right)} \, \mathrm{d}X = \mathrm{d}\mu_{\mathcal{U}}(X),
\end{aligned}$$

where $\mu_{\mathcal{U}}$ is the measure defined by LM, see equation (20). Next let $(\alpha, \beta) \equiv \big((\alpha_C |\, C \in \mathcal{C}), (\beta_S |\, S \in \mathcal{S}) \big) \in \mathbb{R}^{\mathcal{C}} \times \mathbb{R}^{\mathcal{S}}$ satisfy the conditions in (21). Then we have

$$\begin{aligned}
(22)\qquad & H(\alpha, \beta, X) \\
&= \frac{\prod \left(|X_C|^{\alpha_C} \,\big|\, C \in \mathcal{C} \right)}{\prod \left(|X_S|^{\nu(S)\beta_S} \,\big|\, S \in \mathcal{S} \right)} = \frac{\prod \left(|X_{C_j}|^{\alpha_{C_j}} \,\big|\, j = 1, \cdots, \mathcal{C} \right)}{\prod \left(|X_{S_j}|^{\beta_{S_j}} \,\big|\, j = 2, \cdots, \mathcal{C} \right)} \\
&= |X_{C_1}|^{\alpha_1} \prod \left(|X_{R_j\bullet}|^{\alpha_{C_j}} \,\big|\, j = 2, \cdots, \mathcal{C} \right) \prod \left(|X_{S_j}|^{\alpha_{C_j} - \beta_{S_j}} \,\big|\, j = 2, \cdots, \mathcal{C} \right) \\
&= |X_{C_1}|^{\alpha_1} \prod \left(|X_{R_j\bullet}|^{\alpha_{C_j}} \,\big|\, j = 2, \cdots, \mathcal{C} \right) \\
&\quad \times \prod \left(\prod \left(|X_S|^{\alpha_{C_j} - \beta_S} \,\big|\, j \in J_P(S) \right) \Big|\, S \in \mathcal{S} \right) \\
&= |X_{C_1}|^{\alpha_1} \, |X_{S_2}|^{\delta_2} \prod \left(|X_{R_j\bullet}|^{\alpha_{C_j}} \,\big|\, j = 2, \cdots, \mathcal{C} \right) \\
&\quad \times \prod \left(|X_S|^{\sum (\alpha_{C_j} \,|\, j \in J_P(S)) - \nu(S)\beta_S} \,\big|\, S \in \mathcal{S} \setminus S_2 \right)
\end{aligned}$$

$$\times\, |X_{C_1\setminus S_2\bullet}|^{\alpha_1}\, |X_{S_2}|^{\alpha_1+\delta_2} \prod \big(|X_{R_j\bullet}|^{\alpha_{C_j}} \,\big|\, j=2,\cdots,\mathcal{C}\big)\cdot 1$$
$$= \prod \big(|X_{B\bullet}|^{\lambda_B} \,\big|\, B\in V/{\sim}\big)$$

with

$$\lambda_B = \begin{cases} \alpha_1+\delta_2 & \text{for } B=S_2\\ \alpha_{C_1} & \text{for } B=C_1\setminus S_2\\ \alpha_{C_j} & \text{for } B=R_j,\ j=2,\cdots,\mathcal{C}\end{cases}. \tag{23}$$

Furthermore we have

$$\frac{[B]+\langle B\rangle-1}{2} = \begin{cases} \frac{S_2-1}{2} & \text{for } B=S_2\\ \frac{C_1-1}{2} & \text{for } B=C_1\setminus S_2\\ \frac{C_j-1}{2} & \text{for } B=R_j,\ j=2,\cdots,\mathcal{C}\end{cases}.$$

Thus LM's conditions in (21) are equivalent to our condition (17). This establishes the family of Wishart distributions $W_{\mathbf{P}(\mathcal{U}),\alpha,\beta,\sigma}$ of type I, $\big((\alpha,\beta),\sigma\big)\in\mathcal{A}_P\times\mathbf{P}(\mathcal{U})$, to be the family of generalized Riesz distributions $\mathbb{R}_{\Sigma,\lambda}$, $(\Sigma,\lambda)\in\mathbf{P}(\mathcal{U})\times \bigtimes\Big(\big]\frac{[B]+\langle B\rangle-1}{2},\infty\big[\,\Big|\, B\in V/{\sim}\Big)$ with respect to $\mathcal{V}_P$. The one-to-one correspondence between the two parameter sets is given by (23) and by $\Sigma^{-\lambda}=\sigma^{-1}$ (the latter in LM's notation). In particular, $\mathcal{A}_P$ has dimension $V/{\sim}=\mathcal{C}+1$. In general, however, there are representations of $\mathcal{U}$ as an AMG that are not induced by any perfect ordering of the cliques of $\mathcal{U}$. As a consequence our notion of generalized Riesz distributions is more general than LM's notion of Wishart distributions of type I.

Note that our generalized Riesz distributions satisfy $\mathbb{E}\,(\mathbb{R}_{\Sigma,\lambda})=\Sigma$ while the expectation of LM's Wishart distributions of type I is a non-trivial function of the distribution's parameters. From our point of view only the natural parameterization or the expectation parameterization should be used.

Note also that LM's functions $H(\alpha,\beta,\cdot)$ are parameterized by (α,β) in the $(\mathcal{C}+\mathcal{S})$-dimensional space $\mathbb{R}^{\mathcal{C}+\mathcal{S}}$, whereas the actual sets of interest $\mathcal{A}_P$ are of dimension $\mathcal{C}+1$. By comparison our functions $S\mapsto\prod\big(|S_{B\bullet}|^{\lambda_B}\,\big|\,B\in V/{\sim}\big)$, paralleling the role of LM's functions $H(\alpha,\beta,\cdot)$, are parameterized by $\lambda\equiv(\lambda_B\,|\,B\in V/{\sim})$ in the $(V/{\sim})$-dimensional space $\mathbb{R}^{V/\sim}$, with the set of interest $\bigtimes\big(\big]\frac{[B]+\langle B\rangle-1}{2},\infty\big[\,\big|\,B\in V/{\sim}\big)$ having the same dimension.

From our point of view the following natural question arises: *Does there exist an intrinsic (canonical) choice of a representation of $\mathcal{U}$ as an AMG?*—The construction of this AMG (appropriately denoted by $\mathcal{V}_{\mathcal{U}}$) should not depend on any kind of arbitrary choice, as for example an arbitrary perfect ordering of the cliques of $\mathcal{U}$ as in LM. The answer to our question is yes, and it will be given in the following two sections.

13. An intrinsic representation of $\mathcal{U}$

Let $\mathcal{V}=(V,F)$ be an AMG whose box graphs are all DUGs. For all $u,v\in V$ with $u\neq v$ we define

$$v\prec_{\mathcal{V}} u \qquad\text{if}\qquad \{u\}\cup\mathrm{nb}_{\mathcal{V}}(u)\ \subset\ \{v\}\cup\mathrm{nb}_{\mathcal{V}}(v).$$

This yields a partial ordering $\prec_{\mathcal{V}}$ of V induced by $\mathcal{V}$ in a natural way. Note that $\prec_{\mathcal{V}}$ is determined solely by the *undirected part*[1] *of* $\mathcal{V}$, the undirected graph $\mathcal{V}^{\mathrm{u}} := (V, F^{\mathrm{u}})$ with $F^{\mathrm{u}} := \big\{(u,v) \in F \big| (v,u) \in F\big\}$. Note furthermore that $u \prec_{\mathcal{V}} v$ implies $u - v$ in $\mathcal{V}$. The relation $\prec_{\mathcal{V}}$ is empty if and only if $\mathcal{V}^{\mathrm{u}}$ is a disjoint union of complete undirected graphs. Finally, if $\mathcal{V}^{\mathrm{u}} = (\mathcal{V}^{\mathrm{u}})_{V_1} \dot\cup (\mathcal{V}^{\mathrm{u}})_{V_2}$ for some decomposition $V = V_1 \dot\cup V_2$, then $\prec_{\mathcal{V}} = \prec_{\mathcal{V}_{V_1}} \dot\cup \prec_{\mathcal{V}_{V_2}}$.

Based on this partial ordering we will now present the construction of an intrinsic representation $\mathcal{V}_{\mathcal{U}}$ of a given DUG $\mathcal{U}$ with vertex set V as an AMG (see Definition 4.1), thus solving the problem posed in the previous section. The correctness proof will be given in Section 14.

Algorithm 13.1. Set $\mathcal{V}_0 := \mathcal{U}$ and $k := 0$.

Step 1: If $(\mathcal{V}_k)^{\mathrm{u}}$ is a disjoint union of complete undirected graphs, then set $\mathcal{V}_{\mathcal{U}} := \mathcal{V}_k$ and stop. Otherwise continue.

Step 2: For all $u, v \in V$ with $v \prec_{\mathcal{V}_k} u$ replace the line $v - u$ in $\mathcal{V}_k$ with the arrow $v \to u$. This yields a mixed graph $\mathcal{V}'_k$.

Step 3: For all $v, v' \in V$ with $v \to v'$ participating in a partially directed cycle in $\mathcal{V}'_k$ convert the arrow $v \to v'$ back to a line $v - v'$. This yields a mixed graph $\mathcal{V}_{k+1}$.

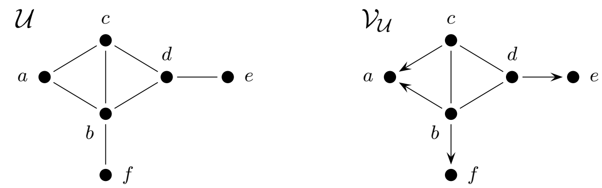
Step 4: Increase k by one and return to Step 1.

We note the following facts: If $\mathcal{U}$ is a disjoint union of complete undirected graphs, then the above algorithm yields $\mathcal{V}_{\mathcal{U}} = \mathcal{U}$; otherwise the relation $\prec_{\mathcal{U}}$ is non-empty and the algorithm will perform Steps 2–4 at least once. All graphs $\mathcal{V}'_k$, $k = 0, 1, 2, \cdots$, and $\mathcal{V}_{k+1}$, $k = 0, 1, 2, \cdots$, have the skeleton $\mathcal{U}$. The mixed graphs $\mathcal{V}'_k$, $k = 0, 1, 2, \cdots$, resulting from Step 2 do not contain any triplexes. The mixed graphs $\mathcal{V}_{k+1}$, $k = 0, 1, 2, \cdots$, resulting from Step 3 are AMGs without triplexes, and all their box graphs are DUGs. In particular, the condition that $(\mathcal{V}_k)^{\mathrm{u}}$ is a disjoint union of complete undirected graphs (see Step 1) is equivalent to requiring all boxes of $\mathcal{V}_k$ to be complete and hence $\mathcal{V}_k$ to be a representation of $\mathcal{U}$ as an AMG. In Step 3, only arrows that have been created in the Step 2 of the same iteration can participate in partially directed cycles; in particular, all arrows created in earlier iterations remain unchanged. Finally, applying Algorithm 13.1 to an unconnected DUG $\mathcal{U}$ is equivalent to applying it separately to all connection components of $\mathcal{U}$.

Example 13.1. Let $\mathcal{U}$ be the *n-vertex chain* $\bullet - \bullet - \cdots - \bullet$ with $V = n \in \mathbb{N}$. Then the intrinsic representation of $\mathcal{U}$ as an AMG found by Algorithm 13.1 is $\bullet \leftarrow \bullet \cdots \leftarrow \bullet - \bullet \to \cdots \bullet \to \bullet$ for even n and $\bullet \leftarrow \bullet \cdots \leftarrow \bullet \to \cdots \bullet \to \bullet$ for odd n. The result is constructed in $\left\lfloor \frac{n-1}{2} \right\rfloor$ iterations, where $\lfloor . \rfloor$ denotes the integer part. For all $k = 0, 1, \cdots, \left\lfloor \frac{n-1}{2} \right\rfloor - 1$ we have $\mathcal{V}'_k = \mathcal{V}_{k+1}$, that is, none of the arrows generated in Step 2 of any iteration are removed in Step 3.

As a second example consider the DUG $\mathcal{U}$ displayed below. Step 2 creates the arrows $b \to a$, $b \to c$, $b \to f$, $c \to a$, and $d \to e$. In the resulting mixed graph $\mathcal{V}'_0$ the arrow $b \to c$ participates in a partially directed cycle. Hence it is converted back to a line in Step 3, thus creating $\mathcal{V}_1 = \mathcal{V}_{\mathcal{U}}$ in one iteration.

[1] Of course one may define the *undirected part* for any mixed graph.



14. Verification of the stepwise construction of $\mathcal{V}_{\mathcal{U}}$

Lemma 14.1. *Let $\mathcal{U} \equiv (V, F)$ be a DUG, and suppose Algorithm 13.1 is applied to $\mathcal{U}$, thus creating the mixed graphs $\mathcal{V}'_k$ and $\mathcal{V}_{k+1}$, $k = 0, 1, 2, \cdots$. Then we have:*

(i) The graphs $\mathcal{V}_{k+1}$, $k = 0, 1, 2, \cdots$, resulting from Step 3 of each iteration are AMGs without triplexes.

(ii) If $(\mathcal{V}_k)^{\mathrm{u}}$ is not a disjoint union of complete undirected graphs for some $k \in \{0, 1, 2, \cdots\}$ (that is, if $\mathcal{V}_k$ does not satisfy the termination condition in Step 1), then the number of arrows in $\mathcal{V}_{k+1}$ is strictly greater than the number of arrows in $\mathcal{V}_k$.

Part (ii) of the above lemma states that each iteration of Algorithm 13.1 produces new arrows. In particular, the number of iterations before termination is trivially bounded by the number of edges in $\mathcal{U}$.

Proof of Lemma 14.1. (i). Assume $u \to v$ to be an arrow participating in a partially directed cycle $u \to v = v_1 \mathrel{\cdots} \cdots \mathrel{\cdots} v_n = u$ in $\mathcal{V}_{k+1}$, where $v_i \mathrel{\cdots} v_{i+1}$, $i = 1, \ldots, n-1$, denotes a line $v_i - v_{i+1}$ or an arrow $v_i \to v_{i+1}$. Then $u \to v$ is also contained in $\mathcal{V}'_k$, and since it is not removed in the transition from $\mathcal{V}'_k$ to $\mathcal{V}_{k+1}$ it does not participate in a partially directed cycle in $\mathcal{V}'_k$. Hence there is $i_0 \in \{1, \ldots, n-1\}$ such that $v_{i_0} \leftarrow v_{i_0+1}$ in $\mathcal{V}'_k$ and $v_{i_0} - v_{i_0+1}$ in $\mathcal{V}_{k+1}$. As a consequence, $v_{i_0} - v_{i_0+1}$ participates in a partially directed cycle $v_{i_0} \to v_{i_0+1} = w_1 \mathrel{\cdots} \cdots \mathrel{\cdots} w_n = v_{i_0}$ in $\mathcal{V}'_k$. In particular, $\mathcal{V}'_k$ contains the cycle $u \to v = v_1 \mathrel{\cdots} \cdots \mathrel{\cdots} v_{i_0-1} \mathrel{\cdots} w_1 \mathrel{\cdots} \cdots \mathrel{\cdots} w_n \mathrel{\cdots} v_{i_0+1} \mathrel{\cdots} \cdots \mathrel{\cdots} v_n = u$. If necessary we may iterate this argument and thus conclude $u \to v$ to participate in a partially directed cycle in $\mathcal{V}'_k$. Hence we have $u - v$ in $\mathcal{V}_{k+1}$, contradicting our assumption. Now the claim follows by induction by the number of iterations of Algorithm 13.1.

(ii). Without loss of generality we may restrict our considerations to an incomplete box B in the AMG $\mathcal{V}_k$ and its box graph $\mathcal{B}$. There exists a replacement of all lines in $\mathcal{B}$ by arrows such that the resulting directed graph $\mathcal{D}$ is acyclic and moral. Let v_{m} be a maximal vertex in $\mathcal{D}$, that is, $v_{\mathrm{m}} \not\to v$, $v \in V$, in $\mathcal{D}$. Since $\mathcal{D}$ is moral the subgraph of $\mathcal{B}$ induced by $\{v_{\mathrm{m}}\} \dot\cup \mathrm{pa}_{\mathcal{D}}(v_{\mathrm{m}}) = \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$ is complete.

Hence we have $\{v\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v) \supseteq \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$ for all $v \in \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$. Since $\mathcal{B}$ is incomplete and connected there are $u \notin \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$ and $v_0 \in \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$ with $u - v_0$. In particular we have $\{v_0\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_0) \supset \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$, which implies $\mathcal{V}'_k$ to contain $v_0 \to v_{\mathrm{m}}$. We will now establish that this arrow will not be removed in the transition from $\mathcal{V}'_k$ to $\mathcal{V}_{k+1}$. To this end we assume $v_0 \to v_{\mathrm{m}}$ to participate in a partially directed cycle $v_0 \to v_{\mathrm{m}} = v_1 \mathrel{\cdots} \cdots \mathrel{\cdots} v_n = v_0$ in $\mathcal{V}'_k$. Then we have $v_2 \in \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$. This entails $\{v_2\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_2) = \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$ and hence $v_{\mathrm{m}} - v_2$ in $\mathcal{V}'_k$. We also have $\{v_0\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_0) \supset \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}})$. Induction by the circle length n yields $\{v_{n-1}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{n-1}) = \{v_{\mathrm{m}}\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_{\mathrm{m}}) \subset \{v_0\} \dot\cup \mathrm{nb}_{\mathcal{B}}(v_0)$. As a consequence we have $v_{n-1} \leftarrow v_0$ in $\mathcal{V}'_k$, contradicting our assumption. $\square$

15. The class of generalized Wishart distributions on $\mathbf{P}(\mathcal{U})$ associated with an undirected decomposable graph

Now we return to the setting from Sections 8–11.

Definition 15.1. Let $\mathcal{V}_\mathcal{U}$ be the intrinsic representation of the DUG $\mathcal{U}$ as an AMG obtained from Algorithm 13.1. Then the Riesz distribution $\mathrm{R}_{\Delta,\lambda}$ on $\mathbf{P}(\mathcal{U})$ with respect to $\mathcal{V}_\mathcal{U}$, with shape parameter $\lambda \equiv (\lambda_B |\, B \in V/{\sim})$ and natural parameter Δ is called the *generalized Wishart distribution* $\mathrm{W}_{\Delta,\lambda}$ *on* $\mathbf{P}(\mathcal{U})$ with *shape parameter* λ and *natural parameter* Δ.

Likewise, the generalized Riesz distribution $\mathbb{R}_{\Sigma,\lambda}$ on $\mathbf{P}(\mathcal{U})$ with respect to $\mathcal{V}_\mathcal{U}$, with shape parameter $\lambda \equiv (\lambda_B |\, B \in V/{\sim})$ and expectation parameter Σ is called the *generalized Wishart distribution* $\mathbb{W}_{\Sigma,\lambda}$ *on* $\mathbf{P}(\mathcal{U})$ with *shape parameter* λ and *expectation parameter* Σ.

The *generalized Wishart model* (with fixed shape parameter λ) in its expectation parameterization is then $\big(\mathbb{W}_{\Sigma,\lambda} \in \mathcal{P}(\mathbf{P}(\mathcal{U}))\big|\, \Sigma \in \mathbf{P}(\mathcal{U})\big)$.

Remark 15.1. Due to $\mathcal{V}_{\mathcal{U}_1 \dot\cup \mathcal{U}_2} = \mathcal{V}_{\mathcal{U}_1} \dot\cup \mathcal{V}_{\mathcal{U}_2}$ Proposition 11.2 remains true when the name Riesz is replaced by the name Wishart. Similarly, the convolution property stated in Proposition 11.3 also holds for generalized Wishart distributions in the sense of Definition 15.1. By contrast, decomposition and marginalization of generalized Wishart distributions do not yield generalized Wishart distributions in general. In this sense, Propositions 10.2 and 11.1 do not carry over to generalized Wishart distributions.

16. The connection to the work by Andersson and Wojnar (2004a,b), Wishart distributions on homogeneous cones.

Now we briefly discuss the present work's relation to the work by Andersson & Wojnar (2004a) and Andersson & Wojnar (2004b), abbreviated as AWa and AWb in the following. Despite its generality the work in AWa,b overlaps with the present work in a non-trivial manner. Given a homogeneous cone C AWa,b define the class of Wishart distributions on C parameterized by its expectation $\Sigma \in C$ and its shape parameter λ, where λ is a finite family of positive scalars as in the present work. In fact, AWa,b indirectly also find Riesz distributions related to the cone C. A self-contained special and easier case of AWa,b can be found in Andersson, Letac & Massam (2008). This special case contains the overlap between the present work and AWa,b lined out and gives further details.

Lemma 16.1. *Let $\mathcal{U} \equiv (V, F)$ be a DUG. Then the following two conditions are equivalent:*

(i) The graph $\mathcal{U}$ does not contain the four chain $\bullet - \bullet - \bullet - \bullet$ as an induced subgraph.
(ii) For all $u, v \in V$ with $u - v$ in $\mathcal{U}$ one of the following three relations holds:
  (a) $\{u\} \dot\cup \mathrm{nb}(u) = \{v\} \dot\cup \mathrm{nb}(v)$,
  (b) $\{u\} \dot\cup \mathrm{nb}(u) \subset \{v\} \dot\cup \mathrm{nb}(v)$, and
  (c) $\{u\} \dot\cup \mathrm{nb}(u) \supset \{v\} \dot\cup \mathrm{nb}(v)$.

Proof. If $\mathcal{U}$ does contain a four chain induced subgraph, then the two central vertices of the four chain do not satisfy (ii). This proves (ii) $\Rightarrow$ (i). Conversely assume (i) and assume $u - v$ not to satisfy (ii). Then there exists $x \in \{u\} \dot\cup \mathrm{nb}(u)$ with

$x \notin \{v\} \dot\cup \mathrm{nb}(v)$ and $y \in \{v\} \dot\cup \mathrm{nb}(v)$ with $y \notin \{u\} \dot\cup \mathrm{nb}(u)$. Clearly x, y, u, v are pairwise distinct. If x and y are not connected, then (i) is violated; otherwise we have a four circle $x - u - v - y - x$ without diagonals, contradicting the decomposability of $\mathcal{U}$. □

A DUG is called *transitive* if it satisfies the equivalent conditions in Lemma 16.1. This definition is consistent with Letac & Massam's (2007) terminology. The term *transitive* is justified since $\mathbf{P}(\mathcal{U})$ is a homogeneous convex cone when $\mathcal{U}$ is transitive, cf. Letac & Massam (2007, Section 2.2) or Andersson & Wojnar (2004a, Example 6.4).

Corollary 16.1. *Let $\mathcal{U}$ be a transitive and incomplete DUG. Then Algorithm 13.1 applied to $\mathcal{U}$ terminates after one iteration, with $\mathcal{V}_0' = \mathcal{V}_1 = \mathcal{V}_{\mathcal{U}}$. In this case the graph of boxes $\mathcal{V}_{\mathcal{U}}/\sim$ is a* transitive *ADG and $\mathcal{V}_{\mathcal{U}}$ does not contain the graph $\bullet - \bullet \to \bullet$ as an induced subgraph.*

Proof. Due to the transitivity (characterized by conditions (a), (b), (c) in Lemma 16.1 (ii)) and the incompleteness of $\mathcal{U}$ the mixed graph $\mathcal{V}_0'$ constructed in Step 2 of the first iteration of Algorithm 13.1 contains neither $\bullet - \bullet - \bullet$ nor $\bullet - \bullet \to \bullet$ nor $\bullet \to \bullet \to \bullet$ as induced subgraphs. Furthermore, $\mathcal{V}_0'$ does not contain any partially directed cycles. Hence $\mathcal{V}_0'$ is an AMG without triplexes and with complete boxes. This proves the claim. □

Now suppose $\mathcal{U}$ to be a transitive DUG and thus $\mathbf{P}(\mathcal{U})$ to be a homogeneous cone, cf. Example 6.4 in AWa. Then the class of Wishart distributions on $\mathbf{P}(\mathcal{U})$ defined in AWa,b coincides with our class of generalized Wishart distributions on $\mathbf{P}(\mathcal{U})$; moreover, both families are parameterized in the same way by expectation and shape parameters.

If $\mathcal{U}$ is non-transitive, then our Definition 15.1 of generalized Wishart distributions is a proper extension of the theory in AWa,b to the case of the convex non-homogeneous cone $\mathbf{P}(\mathcal{U})$. The simplest non-transitive DUG $a - b - c - d$ is thus of special interest.

Example 16.1. In this example we will investigate the different Riesz and Wishart distributions associated with the decomposable undirected graph $a - b - c - d$ (the four-vertex chain), denoted by $\mathcal{U}$. Note first

$$\mathbf{P}(\mathcal{U}) = \left\{ \Sigma \equiv \begin{pmatrix} \sigma_{aa} & \sigma_{ab} & 0 & 0 \\ \sigma_{ba} & \sigma_{bb} & \sigma_{bc} & 0 \\ 0 & \sigma_{cb} & \sigma_{cc} & \sigma_{cd} \\ 0 & 0 & \sigma_{dc} & \sigma_{dd} \end{pmatrix} \middle| \begin{pmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ba} & \sigma_{bb} \end{pmatrix} > 0, \begin{pmatrix} \sigma_{bb} & \sigma_{bc} \\ \sigma_{cb} & \sigma_{cc} \end{pmatrix} > 0, \begin{pmatrix} \sigma_{cc} & \sigma_{cd} \\ \sigma_{dc} & \sigma_{dd} \end{pmatrix} > 0 \right\}.$$

The seven different representations are (i): $a \to b \to c \to d$, (ii): $a \leftarrow b \leftarrow c \leftarrow d$, (iii): $a \leftarrow b \to c \to d$, (iv): $a \leftarrow b \leftarrow c \to d$, (v): $a - b \to c \to d$, (vi): $a \leftarrow b \leftarrow c - d$, and the intrinsic representation (vii): $a \leftarrow b - c \to d$, each generating its own class of Riesz distributions on $\mathbf{P}(\mathcal{U})$. With

$$S \equiv \begin{pmatrix} s_{aa} & s_{ab} & 0 & 0 \\ s_{ba} & s_{bb} & s_{bc} & 0 \\ 0 & s_{cb} & s_{cc} & s_{cd} \\ 0 & 0 & s_{dc} & s_{dd} \end{pmatrix} \in \mathbf{P}(\mathcal{U}), \quad S_{\{ab\}} := \begin{pmatrix} s_{aa} & s_{ab} \\ s_{ba} & s_{bb} \end{pmatrix},$$

and similar definitions for $S_{\{bc\}}$, $S_{\{cd\}}$, and Σ we obtain (i): $\mathrm{d}\mathbb{R}_{\Sigma,\lambda}(S) =$

$$\frac{\pi^{-\frac{3}{2}}\,\lambda_a^{\lambda_a}\,\lambda_b^{\lambda_b}\,\lambda_c^{\lambda_c}\,\lambda_d^{\lambda_d}\,s_{aa}^{\lambda_a-\frac{3}{2}}\,s_{b\bullet}^{\lambda_b-\frac{3}{2}}\,s_{bb}^{-\frac{1}{2}}\,s_{c\bullet}^{\lambda_c-\frac{3}{2}}\,s_{cc}^{-\frac{1}{2}}\,s_{d\bullet}^{\lambda_d-\frac{3}{2}}}{\Gamma(\lambda_a)\,\Gamma(\lambda_b-\frac{1}{2})\,\Gamma(\lambda_c-\frac{1}{2})\,\Gamma(\lambda_d-\frac{1}{2})\,\sigma_{aa}^{\lambda_a}\,\sigma_{b\bullet}^{\lambda_b}\,\sigma_{c\bullet}^{\lambda_c}\,\sigma_{d\bullet}^{\lambda_d}}\;\exp\{-\operatorname{tr}(\Sigma^{-\lambda}S)\}\,\mathrm{d}S,$$

where $\lambda_a > 0$, $\lambda_b > \frac{1}{2}$, $\lambda_c > \frac{1}{2}$, $\lambda_d > \frac{1}{2}$, $\lambda \equiv (\lambda_a, \lambda_b, \lambda_c, \lambda_d) \in \mathbb{R}_+^4$, $s_{b\bullet} := s_{bb} - s_{ba}s_{aa}^{-1}s_{ab}$, $s_{c\bullet} := s_{cc} - s_{cb}s_{bb}^{-1}s_{bc}$, $s_{d\bullet} := s_{dd} - s_{dc}s_{cc}^{-1}s_{cd}$, and similar definitions for $\sigma_{b\bullet}$, $\sigma_{c\bullet}$, and $\sigma_{d\bullet}$.

(iii): $\mathrm{d}\mathbb{R}_{\Sigma,\lambda}(S) =$

$$\frac{\pi^{-\frac{3}{2}}\,\lambda_a^{\lambda_a}\,\lambda_b^{\lambda_b}\,\lambda_c^{\lambda_c}\,\lambda_d^{\lambda_d}\,s_{a\bullet}^{\lambda_a-\frac{3}{2}}\,s_{bb}^{\lambda_b-2}\,s_{c\bullet}^{\lambda_c-\frac{3}{2}}\,s_{cc}^{-\frac{1}{2}}\,s_{d\bullet}^{\lambda_d-\frac{3}{2}}}{\Gamma(\lambda_a-\frac{1}{2})\,\Gamma(\lambda_b)\,\Gamma(\lambda_c-\frac{1}{2})\Gamma(\lambda_d-\frac{1}{2})\,\sigma_{a\bullet}^{\lambda_a}\,\sigma_{bb}^{\lambda_b}\,\sigma_{c\bullet}^{\lambda_c}\,\sigma_{d\bullet}^{\lambda_d}}\;\exp\{-\operatorname{tr}(\Sigma^{-\lambda}S)\}\,\mathrm{d}S,$$

where $\lambda_a > \frac{1}{2}$, $\lambda_b > 0$, $\lambda_c > \frac{1}{2}$, $\lambda_d > \frac{1}{2}$, $\lambda \equiv (\lambda_a, \lambda_b, \lambda_c, \lambda_d) \in \mathbb{R}_+^4$, $s_{a\bullet} := s_{aa} - s_{ab}s_{bb}^{-1}s_{ba}$, $s_{c\bullet} := s_{cc} - s_{cb}s_{bb}^{-1}s_{bc}$, $s_{d\bullet} := s_{dd} - s_{dc}s_{cc}^{-1}s_{cd}$, and similar definitions for $\sigma_{a\bullet}$, $\sigma_{c\bullet}$, and $\sigma_{d\bullet}$.

(v): $\mathrm{d}\mathbb{R}_{\Sigma,\lambda}(S) =$

$$\frac{\pi^{-\frac{3}{2}}\,\lambda_{\{ab\}}^{2\lambda_{\{ab\}}}\,\lambda_c^{\lambda_c}\,\lambda_d^{\lambda_d}\,|S_{\{ab\}}|^{\lambda_{\{ab\}}-\frac{3}{2}}\,s_{bb}^{-\frac{1}{2}}\,s_{c\bullet}^{\lambda_c-\frac{3}{2}}\,s_{cc}^{-\frac{1}{2}}\,s_{d\bullet}^{\lambda_d-\frac{3}{2}}}{\Gamma(\lambda_{\{ab\}})\,\Gamma(\lambda_{\{ab\}}-\frac{1}{2})\,\Gamma(\lambda_c-\frac{1}{2})\,\Gamma(\lambda_d-\frac{1}{2})\,|\Sigma_{\{ab\}}|^{\lambda_{\{ab\}}}\,\sigma_{c\bullet}^{\lambda_c}\,\sigma_{d\bullet}^{\lambda_d}}\;\exp\{-\operatorname{tr}(\Sigma^{-\lambda}S)\}\,\mathrm{d}S,$$

where $\lambda_{\{ab\}} > \frac{1}{2}$, $\lambda_c > \frac{1}{2}$, $\lambda_d > \frac{1}{2}$, $\lambda \equiv (\lambda_{\{ab\}}, \lambda_b, \lambda_c) \in \mathbb{R}_+^3$, $s_{c\bullet} := s_{cc} - s_{cb}s_{bb}^{-1}s_{bc}$, $s_{d\bullet} := s_{dd} - s_{dc}s_{cc}^{-1}s_{cd}$, and similar definitions for $\sigma_{c\bullet}$ and $\sigma_{d\bullet}$.

The cases (ii), (iv), and (vi) are analogous to (i), (iii), and (v), respectively, with a, b, c, and d replaced by d, c, b, and a, respectively.

(vii): $\mathrm{d}\mathbb{R}_{\Sigma,\lambda}(S) \equiv \mathrm{d}\mathbb{W}_{\Sigma,\lambda}(S) =$

$$\frac{\pi^{-\frac{3}{2}}\,\lambda_a^{\lambda_a}\lambda_{\{bc\}}^{2\lambda_{\{bc\}}}\lambda_d^{\lambda_d}\,s_{a\bullet}^{\lambda_a-\frac{3}{2}}\,s_{bb}^{-\frac{1}{2}}|S_{\{bc\}}|^{\lambda_{\{bc\}}-\frac{3}{2}}\,s_{cc}^{-\frac{1}{2}}\,s_{d\bullet}^{\lambda_d-\frac{3}{2}}}{\Gamma(\lambda_a-\frac{1}{2})\Gamma(\lambda_{\{bc\}})\Gamma(\lambda_{\{bc\}}-\frac{1}{2})\Gamma(\lambda_d-\frac{1}{2})\,\sigma_{a\bullet}^{\lambda_a}|\Sigma_{\{bc\}}|^{\lambda_{\{bc\}}}\sigma_{d\bullet}^{\lambda_d}}\;\exp\{-\operatorname{tr}(\Sigma^{-\lambda}S)\}\,\mathrm{d}S,$$

where $\lambda_a > \frac{1}{2}$, $\lambda_{\{bc\}} > \frac{1}{2}$, $\lambda_d > \frac{1}{2}$, $\lambda \equiv (\lambda_a, \lambda_{\{bc\}}, \lambda_c) \in \mathbb{R}_+^3$, $s_{a\bullet} := s_{aa} - s_{ab}s_{bb}^{-1}s_{ba}$, $s_{d\bullet} := s_{dd} - s_{dc}s_{cc}^{-1}s_{cd}$, and similar definitions for $\sigma_{a\bullet}$ and $\sigma_{d\bullet}$.

As indicated above case (vii) is the Wishart distribution on $\mathbf{P}(\mathcal{U})$ with shape parameter $\lambda \equiv (\lambda_a, \lambda_{\{bc\}}, \lambda_d) \in \left]\frac{1}{2}, \infty\right[^3$ and expectation $\Sigma \in \mathbf{P}(\mathcal{U})$.

17. Further Research.

As is the case for the general Wishart distributions on homogeneous cones from Andersson & Wojnar (2004b, Section 7), our results open new interesting areas in multivariate analysis for further investigation, in particular likelihood inference in generalized Riesz/Wishart models.

For instance, consider a given representation $\mathcal{V}$ of $\mathcal{U}$ as an AMG and let S_k, $k \in K$, be independent observables from $\mathbf{P}(\mathcal{U})$ with S_k following the generalized Riesz distribution $\mathrm{R}_{\Delta_k,\lambda_k}$, $\Delta_k \in \mathbf{PD}^0(\mathcal{U})$, $k \in K$. The testing problem $\Delta_k = \Delta$, $k \in K$ (that is, all Δ_k, $k \in K$, are identical) can be solved completely, see Crawford (2008). The solution constitutes a generalization of the classical Bartlett test.

The concept of singular classical Wishart distributions is readily extended to generalized Riesz/Wishart distributions on $\mathbf{P}(\mathcal{U})$ and open for further investigation.

Replacing the cone $\mathbf{P}(\mathcal{U})$ with the dual cone $\mathbf{PD}^0(\mathcal{U})$ (of positive definite matrices) one can develop definitions of generalized Riesz/Wishart distributions similar to those of the present paper, see Andersson & Klein (2008). These distributions are related to LM's *Wishart distributions of type II*; more precisely, their distributions are special cases of generalized Riesz distributions on $\mathbf{PD}^0(\mathcal{U})$ but use a different parameterization. Expectations of these new distributions are not easily expressed in terms of the distributions' parameters. Moreover, these families of distributions are, in general, not closed under convolution (cf. Proposition 11.3). This lack of closure under convolution is found even in the special case of LM's Wishart distributions of type II, despite LM's claim "We have parallel results for the type II Wishart" (LM, top of p. 1311, referring to the closure property of families of Wishart distributions of type I, stated in the first equation on p. 1310).

References

Andersson, S. A. & Klein, T. (2008). On Riesz and Wishart distributions associated with decomposable undirected graphs II. In preparation.

Andersson, S. A., Letac, G. & Massam, H. (2008). The variance of the Wishart distributions on homogeneous cones. In preparation.

Andersson, S. A., Madigan, D. & Perlman, M. D. (2001). Alternative Markov properties for chain graphs, *Scandinavian Journal of Statistics* **28**(1): 33–85.

Andersson, S. A. & Perlman, M. D. (1998). Normal linear regression models with recursive graphical Markov structure, *Journal of Multivariate Analysis* **66**: 133–187.

Andersson, S. A. & Wojnar, G. G. (2004a). Wishart distributions on homogeneous cones, *Journal of Theoretical Probability* **17**(4): 781–818.

Andersson, S. A. & Wojnar, G. G. (2004b). The Wishart distributions on homogeneous cones, *Acta et Commentationes Universitatis Tartuensis de Mathematica* **8**: 3–62. Special Volume: Proceedings of the 7th Tartu Conference on Multivariate Statistics, August 7–12, 2003, Tartu, Estonia.

Casalis, M. (1991). Les familles exponentielles à variance quadratique homogène sont des lois de Wishart sur un cône symétrique, *Comptes Rendus de l'Académie des Sciences—Series I—Mathematics* **312**: 537–540.

Casalis, M. & Letac, G. (1996). The Lukacs-Olkin-Rubin characterization of Wishart distributions on symmetric cones, *Annals of Statistics* **24**(2): 763–786.

Crawford, J. (2008). Bartlett's test for equality of multivariate scales in generalized Riesz distributions. In preparation.

Faraut, J. & Korányi, A. (1994). *Analysis on symmetric cones*, Oxford Mathematical Monographs, Clarendon Press, Oxford.

Hassairi, A. & Lajmi, S. (2001). Riesz exponential families on symmetric cones, *Journal of Theoretical Probability* **14**(4): 927–948.

Lauritzen, S. L. (1996). *Graphical Models*, Oxford Statistical Science Series, Clarendon Press, Oxford.

Letac, G. & Massam, H. (1998). Quadratic and inverse regressions for Wishart distributions, *Annals of Statistics* **26**(2): 573–595.

Letac, G. & Massam, H. (2007). Wishart distributions for decomposable graphs, *Annals of Statistics* **35**(3): 1278–1323.

Massam, H. & Neher, E. (1997). On transformations and determinants of Wishart variables on symmetric cones, *Journal of Theoretical Probability* **10**(4): 867–902.

Riesz, M. (1949). L'intégrale de Riemann-Liouville et le problème de Cauchy, *Acta Mathematica* **81**: 1–223.

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population, *Biometrika* **20A**(1/2): 32–52.

Department of Statistics, Indiana University, Bloomington, IN 47405

Institut für Mathematik, Universität Augsburg, Germany

E-mail address: standers@indiana.edu, thomas.klein@math.uni-augsburg.de