Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Multivariate Analysis 97 (2006) 103 – 123

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

The Matsumoto–Yor property and the structure of the Wishart distribution

Hélène Massam[a,*], Jacek Wesołowski[b]

[a]Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ont., Canada M3J1P3
[b]Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, Poland

Received 17 February 2004
Available online 20 January 2005

Abstract

This paper establishes a link between a generalized matrix Matsumoto–Yor (MY) property and the Wishart distribution. This link highlights certain conditional independence properties within blocks of the Wishart and leads to a new characterization of the Wishart distribution similar to the one recently obtained by Geiger and Heckerman but involving independences for only three pairs of block partitionings of the random matrix.

In the process, we obtain two other main results. The first one is an extension of the MY independence property to random matrices of different dimensions. The second result is its converse. It extends previous characterizations of the matrix generalized inverse Gaussian and Wishart seen as a couple of distributions.

We present two proofs for the generalized MY property. The first proof relies on a new version of Herz's identity for Bessel functions of matrix arguments. The second proof uses a representation of the MY property through the structure of the Wishart.

AMS 1991 subject classification: 62H05; 62E10; 62-09

Keywords: Characterization of the Wishart; GIG distribution; Matsumoto–Yor property; Graphical Gaussian models

* Corresponding author.
E-mail address: massamh@mathstat.yorku.ca (H. Massam).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2004.11.008

1. Introduction

In recent years the Wishart distribution and distributions derived from the Wishart have received a lot of attention because of their use in graphical Gaussian models. The essence of graphical models in multivariate analysis is to identify independences and conditional independences between various groups of variables. For Gaussian models, this leads to considering the Wishart distribution and the independences between blocks of the Wishart matrix. In this perspective, Geiger and Heckerman [5] searched for a characterization of the Wishart distribution for 2×2 random matrices. Realizing this is not possible, they obtained instead a characterization of the quasi-Wishart. Letac and Massam [10] gave an alternate proof of their result using the converse of the univariate Matsumoto–Yor (MY) [14] property (henceforth abbreviated as MY property), which will be described below. Pursuing their earlier work, Geiger and Heckerman [6] found a new characterization of the Wishart for matrices of dimensions greater than or equal to three. For the Wishart variate K partitioned into blocks (K_1, K_{12}, K_2), this characterization involves the classical independence properties of $K_{2\cdot 1} = K_2 - K_{21}K_1^{-1}K_{12}$ and (K_1, K_{12}) and of $K_{1\cdot 2} = K_1 - K_{12}K_2^{-1}K_{21}$ and (K_2, K_{12}). Now, these imply the conditional independences of K_1 and $K_{2\cdot 1}$ given K_{12} and of K_2 and $K_{1\cdot 2}$ given K_{21}. The latter call to mind the MY property and it is natural to wonder whether it is possible to find a multivariate version of the MY property with matrices of different dimensions which links the Wishart and its characterization to this new MY property just as it was done for 2×2 matrices in [10]. Accordingly, in this paper we define and study the multivariate MY property for matrix variates of different dimensions and establish the interplay between this MY property and the conditional structure of the Wishart.

Recent works in graphical Gaussian models in a classical or Bayesian framework have shown the importance of the parametrization of the type $K_1, K_1^{-1}K_{12}, K_{2\cdot 1}$ for the precision matrix $K = \Sigma^{-1}$. In this context understanding the role of the MY property in the structure of the Wishart appears essential.

Let us now sketch here a brief history of the development of the MY property, which is relatively recent, and give an outline of the paper.

Since the way of parametrizing the *GIG* and gamma distributions is of some importance here, we fix it in the following way. The generalized inverse Gaussian distribution $GIG(-p, a, b)$ is defined by its density

$$f(x) = \frac{1}{K_p(a,b)} x^{-p-1} e^{-ax-b/x} \mathbf{1}_{(0,\infty)}(x),$$

where either $p > 0, a \geqslant 0, b > 0$, or $p < 0, a > 0, b \geqslant 0$ or $p = 0, a > 0, b > 0$ and $K_p(a,b) = a^{p/2} b^{-p/2} K_p(2\sqrt{ab})$, with $\lambda \mapsto K_k(\lambda)$ being the modified Bessel function of the third kind. The gamma distribution $W(q, c)$ is defined by its Laplace transform

$$L(\theta) = \left(\frac{c}{c-\theta}\right)^q,$$

where $q \geqslant 0$ and $c > 0$. If $q = 0$ the distribution is degenerate at zero, while for $q > 0$ its density has the form

$$g(y) = \frac{c^q}{\Gamma(q)} y^{q-1} e^{-cy} \mathbf{1}_{(0,\infty)}(y).$$

While studying functionals of exponential Brownian motion, Matsumoto and Yor [14] considered the transformation ψ that takes $(x, y) \in (0, \infty)^2$ into $(0, \infty)^2$, where

$$\psi(x, y) = \left((x + y)^{-1}, x^{-1} - (x + y)^{-1}\right). \tag{1.1}$$

They observed that if $(X, Y) \sim GIG(-p, a, b) \otimes W(p, a)$ then $(U, V) = \psi(X, Y) \sim GIG(-p, b, a) \otimes W(p, b)$ for positive p, a, b (see also [15]).

Letac and Wesołowski [11] considered the transformation ψ for X and Y positive definite random matrices of the same dimension and proved that the MY property still holds. Moreover under certain smoothness conditions they proved the converse, that is the following characterization, if X and Y are independent $r \times r$ positive definite matrices and $U = (X + Y)^{-1}$ and $V = X^{-1} - (X + Y)^{-1}$ are also independent then X and Y follow a matrix variate *GIG* (see (2.1) below) and Wishart distribution, respectively. They also proved that the smoothness conditions are no longer necessary in the univariate case. (Multivariate versions of the MY property driven by a graphical structure of trees and related characterizations of the *GIG* and gamma distributions have been developed recently in [13].)

Here we consider a more general situation: X and Y are positive definite random matrices of dimensions $r \times r$ and $s \times s$, respectively, z is a constant $s \times r$ matrix of full rank. We define $U = (zXz^t + Y)^{-1}$ and $V = X^{-1} - z^t(zXz^t + Y)^{-1}z$. First, in Section 3, we show that if X is a matrix variate *GIG* and Y is a Wishart with suitably chosen parameters and they are independent, then U and V are independent and respectively follow a matrix variate *GIG* and Wishart distribution with specified parameters. We give two proofs for this result. The first one relies on a new version of Herz's identity for Bessel functions of matrix arguments; the second one on a conditional independence property of the Wishart and a result of Butler [2]. In Section 4, we give the characterization of the matrix variate *GIG* and Wishart distribution through the independence of X and Y, and U and V, under certain smoothness assumptions. This is an extension of the characterizations obtained in [11,17] which were concerned only with random matrices of the same dimensions and $z = I_r$. Our proof makes use of a functional equation result given in [17].

In Section 5, we give a characterization of the Wishart distribution closely related to that given by Geiger and Heckerman [6]. They showed that for $n \geqslant 3$, an $n \times n$ random matrix K with absolutely continuous distribution is Wishart if and only if $K_{2\cdot 1}$ and (K_{12}, K_2) are independent for every block partitioning K_1, K_{12}, K_2 of K. Our characterization of the Wishart distribution requires independences of that type for only three pairs of block partitionings of K, such that K_1 is a 1×1 matrix. The proof relies on the characterization of the *GIG* and Wishart given in Section 4 and on the link between the MY property and the Wishart already established in Section 3 (see the second proof of Theorem 3.1).

In the next section we set the notation adopted throughout the paper and give some preliminary results to be used later on.

2. Preliminaries

We will briefly address the following: the definitions of the Wishart and the matrix variate *GIG* distributions, the basics of the Cholesky decomposition, the definition and some elementary properties of the mapping ψ_z which is a natural extension to matrix arguments of the mapping ψ defined in the introduction, a formula for the covariance of aX and bX when X is a normal matrix variate and a and b are given non-random matrices and finally a useful property of conditional independence.

Let $\mathcal{V}_n$ be the Euclidean space of $n \times n$ real symmetric matrices equipped with the inner product $\langle a, b\rangle = \operatorname{tr}(ab)$. Let dx denote the Lebesgue measure on $\mathcal{V}_n$ assigning the unit mass to the unit cube. Let $\mathcal{V}_n^+$ denote the cone of positive definite matrices in $\mathcal{V}_n$ and let $\overline{\mathcal{V}}_n^+$ denote its closure. For $x \in \mathcal{V}_n$ let $|x|$ denote the determinant of x.

Let $c \in \mathcal{V}_n^+$ and $q \in \Lambda_n = \{0, \frac{1}{2}, \frac{2}{2}, \ldots, \frac{n-1}{2}\} \cup (\frac{n-1}{2}, \infty)$. Then the random matrix Y taking its values in $\overline{\mathcal{V}}_n^+$ is said to follow the Wishart $W_n(q, c)$ distribution if its Laplace transform is

$$L_Y(\theta) = \frac{|c|^q}{|c-\theta|^q}, \quad c - \theta \in \mathcal{V}_n^+.$$

Note that here, we conveniently parameterize the Wishart matrix Y by taking $qc^{-1} = E(Y)$. It is well known that the above formula is the Laplace transform of a probability measure if and only if $c \in \mathcal{V}_n^+$ and $q \in \Lambda_n$. The set Λ_n is called a Gindikin set (see [3,7]). When $q > \frac{n-1}{2}$, that is when Y takes its values in $\mathcal{V}_n^+$, this distribution has density of the form

$$f_Y(y) = \frac{|c|^q}{\Gamma_n(q)} |y|^{q-\frac{n+1}{2}} \exp(-\langle c, y\rangle) \mathbf{1}_{\mathcal{V}_n^+}(y),$$

where Γ_n is the multivariate Gamma function, see [16]. When $q \in \Lambda_n$ and $q < \frac{n-1}{2}$ the distribution is singular and is concentrated on the boundary of $\bar{\mathcal{V}}_n^+$. In the special case $q = 0$, it is the Dirac measure concentrated at the zero matrix.

A random matrix X, taking its values in $\mathcal{V}_n^+$, is said to follow the $GIG_n(-p, a, b)$ distribution if it has density of the form

$$f_X(x) = \frac{1}{K_p^{(n)}(a, b)} |x|^{-p-\frac{n+1}{2}} \exp\left(-\langle a, x\rangle - \langle b, x^{-1}\rangle\right) \mathbf{1}_{\mathcal{V}_n^+}(x), \tag{2.1}$$

where $K_p^{(n)}$ is the matrix variate modified Bessel function of the third kind, see [8]. (The superscript (n) indicating the dimension of the matrix arguments will be omitted in the sequel, since the dimension will always be obvious from the context.) Letac [9] has observed that the $GIG_n(-p, a, b)$ is well defined iff p, a, b satisfy one of the following three conditions:

1. $a, b \in \mathcal{V}_n^+$ and $p \in \mathbb{R}$,
2. $a \in \bar{\mathcal{V}}_n^+$ with $\operatorname{rank}(a) = m \in \{0, 1, \ldots, n-1\}$, $b \in \mathcal{V}_n^+$ and $p > \frac{n-m-1}{2}$,
3. $a \in \mathcal{V}_n^+$, $b \in \bar{\mathcal{V}}_n^+$ with $\operatorname{rank}(b) = m \in \{0, 1, \ldots, n-1\}$ and $p < -\frac{n-m-1}{2}$.

This extends earlier definitions of the matrix variate *GIG* as given in [1] or [2].

It is immediate to see that the *GIG* distribution has the following property

$$\text{if} \quad X \sim GIG_n(-p, a, b) \quad \text{then} \quad X^{-1} \sim GIG_n(p, b, a), \tag{2.2}$$

which can be rephrased in terms of Bessel functions as

$$K_{-p}(a, b) = K_p(b, a). \tag{2.3}$$

For $X \sim GIG_n(p, a, b)$ and for $a + \theta$ and $b + \sigma$ satisfying one of conditions 1–3 above, we have the following obvious identity

$$E(\exp(\langle\theta, X\rangle + \langle\sigma, X^{-1}\rangle)) = \frac{K_p(a + \theta, b + \sigma)}{K_p(a, b)}. \tag{2.4}$$

For any matrix $k \in \mathcal{V}_n$ consider the following block partitioning

$$k = \begin{pmatrix} k_1 & k_{12} \\ k_{21} & k_2 \end{pmatrix}$$

with k_1 of dimension $r \times r$, $k_{12} = k_{21}^t$ of dimension $r \times s$, k_2 of dimension $s \times s$ and $r + s = n$.

For $k \in \overline{\mathcal{V}}_n^+$ and $k_1 \in \mathcal{V}_r^+$ we consider the following block Cholesky (or Frobenius) decomposition

$$k = \tau_k \begin{pmatrix} k_1 & 0 \\ 0 & k_{2\cdot 1} \end{pmatrix} \tau_k^t \ , \quad \text{where} \quad \tau_k = \begin{pmatrix} I_r & 0 \\ k_{21}k_1^{-1} & I_s \end{pmatrix} \tag{2.5}$$

and $k_{2\cdot 1} = k_2 - k_{21}k_1^{-1}k_{12}$ and I_r and I_s are unit matrices of dimensions r and s, respectively. We note that the rank of k is equal to the rank of $\begin{pmatrix} k_1 & 0 \\ 0 & k_{2\cdot 1} \end{pmatrix}$. So, $k \in \mathcal{V}_n^+$ if and only if $k_{2\cdot 1} \in \mathcal{V}_s^+$. From (2.5) we see that

$$|k| = |k_1||k_{2\cdot 1}|. \tag{2.6}$$

In a dual way, for $k \in \overline{\mathcal{V}}_n^+$ and $k_2 \in \mathcal{V}_s^+$ we consider the following block Cholesky (or Frobenius) decomposition

$$k = \rho_k^t \begin{pmatrix} k_{1\cdot 2} & 0 \\ 0 & k_2 \end{pmatrix} \rho_k \ , \quad \text{where} \quad \rho_k = \begin{pmatrix} I_r & 0 \\ k_2^{-1}k_{21} & I_s \end{pmatrix} \tag{2.7}$$

and $k_{1\cdot 2} = k_1 - k_{12}k_2^{-1}k_{21}$. We note that the rank of k is equal to the rank of $\begin{pmatrix} k_{1\cdot 2} & 0 \\ 0 & k_2 \end{pmatrix}$. So, $k \in \mathcal{V}_n^+$ if and only if $k_{1\cdot 2} \in \mathcal{V}_r^+$. From (2.7) we see that

$$|k| = |k_2||k_{1\cdot 2}|. \tag{2.8}$$

If $k \in \mathcal{V}_n^+$ then from (2.6) and (2.8) it follows that

$$|k| = |k_1||k_{2\cdot 1}| = |k_2||k_{1\cdot 2}|. \tag{2.9}$$

For any $s \times r$ real matrix z of full rank, we will denote by $\mathbf{P}(z)$ the linear mapping

$$x \in \mathcal{V}_r \mapsto \mathbf{P}(z)x = zxz^t \in \mathcal{V}_s.$$

As mentioned in the introduction we generalize the transformation ψ defined in (1.1). To do so we need the following two lemmas.

Lemma 2.1. *Let z be an $s \times r$ real matrix of full rank. For $(x, y) \in \mathcal{V}_r^+ \times \overline{\mathcal{V}}_s^+$ and such that $\mathbf{P}(z)x + y$ is positive definite we define ψ_z as follows*:

$$\begin{aligned}\psi_z(x, y) &= (u(x, y), v(x, y)) \\ &= \left((\mathbf{P}(z)x + y)^{-1},\ x^{-1} - \mathbf{P}(z^t)(\mathbf{P}(z)x + y)^{-1}\right).\end{aligned} \tag{2.10}$$

Then (u, v) belongs to $\mathcal{V}_s^+ \times \overline{\mathcal{V}}_r^+$.

Moreover, $y \in \mathcal{V}_s^+$ if and only if $v \in \mathcal{V}_r^+$, which is equivalent to $y \in \overline{\mathcal{V}}_s^+ \setminus \mathcal{V}_s^+$ if and only if $v \in \overline{\mathcal{V}}_r^+ \setminus \mathcal{V}_r^+$.

Proof. Clearly $(u, v) \in \mathcal{V}_s^+ \times \mathcal{V}_r$. For $(x, y) \in \mathcal{V}_r^+ \times \overline{\mathcal{V}}_s^+$ we define

$$k_1 = x^{-1}\ , \quad k_{21} = z = k_{12}^t\ , \quad k_2 = \mathbf{P}(z)x + y$$

so that

$$x = k_1^{-1}\ , \quad y = k_{2\cdot 1}\ , \quad u = k_2^{-1}\ , \quad v = k_{1\cdot 2}.$$

We consider the symmetric $(r + s) \times (r + s)$ matrix k

$$k = \begin{pmatrix} k_1 & k_{12} \\ k_{21} & k_2 \end{pmatrix}.$$

From (2.5) and (2.7) it follows that

$$k = \tau_k \begin{pmatrix} x^{-1} & 0 \\ 0 & y \end{pmatrix} \tau_k^t = \rho_k^t \begin{pmatrix} v & 0 \\ 0 & u^{-1} \end{pmatrix} \rho_k.$$

Since $y \in \overline{\mathcal{V}}_s^+$ then k is positive semi-definite and so is v. Moreover since x and u are positive definite it follows that $y \in \mathcal{V}_s^+$ if and only if $v \in \mathcal{V}_r^+$. □

Lemma 2.2. *The mapping ψ_z is a bijection from $\mathcal{V}_r^+ \times \mathcal{V}_s^+$ onto $\mathcal{V}_s^+ \times \mathcal{V}_r^+$. The absolute value of the Jacobian of ψ_z^{-1} is equal to*

$$|J| = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| = |u|^{-(s+1)}|v + \mathbf{P}(z^t)u|^{-(r+1)}. \tag{2.11}$$

Proof. From the previous lemma we know that ψ_z is a mapping from $\mathcal{V}_r^+ \times \mathcal{V}_s^+$ into $\mathcal{V}_s^+ \times \mathcal{V}_r^+$. Conversely, ψ_{z^t} is a mapping from $\mathcal{V}_s^+ \times \mathcal{V}_r^+$ into $\mathcal{V}_r^+ \times \mathcal{V}_s^+$ and therefore ψ_z is a bijection and $\psi_z^{-1} = \psi_{z^t}$.

Let us now compute the Jacobian of ψ_z^{-1}. We decompose ψ_z as $\psi_z = \phi_2 \circ \phi_1$, where $\phi_1 : \mathcal{V}_r^+ \times \mathcal{V}_s^+ \to \Delta \subset \mathcal{V}_r^+ \times \mathcal{V}_s^+$ with Δ being the image of $\mathcal{V}_r^+ \times \mathcal{V}_s^+$ by ϕ_1, is defined by

$$\phi_1(x, y) = (\alpha(x, y), \beta(x, y)) = (x, \mathbf{P}(z)x + y)$$

and $\phi_2 : \Delta \to \mathcal{V}_s^+ \times \mathcal{V}_r^+$ is defined by

$$\phi_2(\alpha, \beta) = (u(\alpha, \beta), v(\alpha, \beta)) = (\beta^{-1}, \alpha^{-1} - \mathbf{P}(z^t)(\beta^{-1})).$$

Let us note that both ϕ_1 and ϕ_2 are bijections with respective Jacobians

$$\frac{\partial(\alpha, \beta)}{\partial(x, y)} = J_1 = \mathrm{Det}\begin{pmatrix} id_{\mathcal{V}_r^+} & * \\ 0 & id_{\mathcal{V}_s^+} \end{pmatrix}$$

and

$$\frac{\partial(u, v)}{\partial(\alpha, \beta)} = J_2 = \mathrm{Det}\begin{pmatrix} -\mathbf{P}(\beta^{-1}) & * \\ 0 & -P(\alpha^{-1}) \end{pmatrix},$$

where the parts of J_1 and J_2 not needed in our calculations have been denoted by * and Det denotes the determinant of operators. Therefore

$$\begin{aligned} \frac{\partial(u, v)}{\partial(x, y)} &= J_1 J_2 = \mathrm{Det}(\mathbf{P}(\beta^{-1}))\mathrm{Det}(\mathbf{P}(\alpha^{-1})) \\ &= |\beta|^{-(s+1)}|\alpha|^{-(r+1)} = |x|^{-(r+1)}|\mathbf{P}(z)x + y|^{-(s+1)}. \end{aligned}$$

The equality before last follows from Theorem 2.1.7 in [16]. Since $x = (\mathbf{P}(z^t)u + v)^{-1}$ and $\mathbf{P}(z)x + y = u^{-1}$, (2.11) follows immediately. $\square$

Let $\mathcal{M}_{m,n}$ be the space of real $m \times n$ matrices with inner product $\langle u, v\rangle = \mathrm{tr}\,(u^t v)$.

Let us recall that given two matrix variates X and Y taking their values in $\mathcal{M}_{m,n}$, the covariance operator is defined as the unique symmetric bilinear form $\mathrm{Cov}(X, Y) : \mathcal{M}_{m,n} \times \mathcal{M}_{m,n} \to \mathbb{R}$ such that for all $u, v \in \mathcal{M}_{m,n}$

$$\langle u, \mathrm{Cov}(X, Y)v\rangle = \mathrm{cov}(\langle u, X\rangle, \langle v, Y\rangle).$$

Given $a \in \mathcal{M}_{m,m}$ and $b \in \mathcal{M}_{n,n}$, the Kronecker product $a \otimes b$ of a and b is the bilinear form on $\mathcal{M}_{m,n} \times \mathcal{M}_{m,n}$ defined by $(u, v) \mapsto (a \otimes b)(u, v) = \langle u, avb^t\rangle$.

When X in an $m \times n$ normal matrix variate, $\mathrm{Cov}(X) = \sigma_1 \otimes \sigma_2$ for some $\sigma_1 \in \mathcal{V}_m^+$ and $\sigma_2 \in \mathcal{V}_n^+$. If $a, b \in \mathcal{M}_{m,m}$ are constant, then clearly

$$\mathrm{Cov}(aX,\ bX) = (a\sigma_1 b^t) \otimes \sigma_2. \tag{2.12}$$

Finally let us recall a conditional independence property that we will often use here. For any random variates X, Y, Z

$$\{\ (X, Y) \perp\!\!\!\perp Z\ \} \Leftrightarrow \{\ (X \perp\!\!\!\perp Z)|Y \quad \text{and} \quad Y \perp\!\!\!\perp Z\ \}. \tag{2.13}$$

3. The Matsumoto–Yor property for matrix variates of different dimensions

In this section, we give a further extension of this property to the case where X and Y are random matrices of not necessarily the same dimension using the mapping ψ_z as defined in (2.10). In the special case when X and Y have the same dimension and z is the identity matrix our result reduces to the MY property given by Letac and Wesołowski [11]. The direct MY property for random matrices of different dimensions is as follows.

Theorem 3.1. *Let X and Y be two independent random matrices with GIG and Wishart distributions*

$$X \sim GIG_r(-p, \mathbf{P}(z^t)a, b) \quad \text{and} \quad Y \sim W_s(q, a) \tag{3.1}$$

such that $q \in \Lambda_s$ *and* $p = q + \frac{r-s}{2} \in \Lambda_r$. *Let z be a given constant* $s \times r$ *matrix of full rank. Then* $\mathbf{P}(z)X + Y$ *is positive definite a.s. Moreover U and V defined as*

$$U = (\mathbf{P}(z)X + Y)^{-1} \quad \text{and} \quad V = X^{-1} - \mathbf{P}(z^t)(\mathbf{P}(z)X + Y)^{-1}$$

are independent with GIG and Wishart distributions

$$U \sim GIG_s(-q, \mathbf{P}(z)b, a) \quad \text{and} \quad V \sim W_r(p, b). \tag{3.2}$$

We are going to present two proofs of this result. The first proof uses Laplace transform techniques, while the second relies on the normal representation of the Wishart. We think that both methods are interesting. The first one relies on a new identity for Bessel functions of matrix variates of different dimensions which extends an earlier result by Hertz [8] and a later result by Letac and Wesołowski [11]. The second one relates the MY property to the conditional structure of the Wishart matrix.

3.1. First proof: using a Bessel functions identity

Our basic tool is a new identity for Bessel functions given in the following proposition. Its proof is deferred to the Appendix A.

Proposition 3.1. *Let* $a \in \mathcal{V}_s^+$, $b \in \mathcal{V}_r^+$ *and let z be an* $s \times r$ *matrix of full rank. Let* $p, q > -1/2$ *and* $q = p - \frac{r-s}{2}$. *Then*

$$\Gamma_s(q)|b|^p K_p(b, \mathbf{P}(z^t)a) = \Gamma_r(p)|a|^q K_q(a, \mathbf{P}(z)b). \tag{3.3}$$

We note that (3.1) makes sense for p or q in $(-1/2, 0)$ with, in this case, the following convention

$$\frac{\Gamma_r(p)}{\Gamma_s(q)} = \begin{cases} \Gamma_{r-s}(p)\pi^{\frac{s(r-s)}{2}} & \text{if } s < r, \text{ (which implies } p > 0), \\ \left[\Gamma_{s-r}(q)\pi^{\frac{r(r-s)}{2}}\right]^{-1} & \text{if } s > r, \text{ (which implies } q > 0), \\ 1 & \text{if } r = s, \text{ (which implies } p = q). \end{cases}$$

For arbitrary r and s and $p > \frac{r-1}{2}$, identity (3.3) was given in [8]. For $r = s$, $z = I_r$ and arbitrary $p = q$ it was given in [11] as

$$|b|^p K_p(b, a) = |a|^p K_p(a, b). \tag{3.4}$$

They used it in the proof of the MY property for matrices of the same dimensions. We will follow a similar pattern.

Proof of Theorem 3.1. Let us first assume that $s \leqslant r$. Then $\mathbf{P}(z)X + Y$ is positive definite a.s. and so U is well defined. We therefore consider the transform

$$M_{(U,V)}(\theta, \sigma) = E\left(e^{\langle\theta, \mathbf{P}(z^t)U+V\rangle + \langle\sigma, U^{-1}\rangle}\right)$$

for $\theta \in \mathcal{V}_r^+$ and $\sigma \in \mathcal{V}_s^+$. It uniquely determines the distribution of (U, V) since it is the Laplace transform of the distribution of $(\mathbf{P}(z^t)U + V, U^{-1})$. By (3.1), the independence assumption, formulas (2.4) and (2.3) we have

$$\begin{aligned} M_{(U,V)}(\theta, \sigma) &= E\left(e^{\langle\theta, X^{-1}\rangle + \langle\sigma, \mathbf{P}(z)X+Y\rangle}\right) = E\left(e^{\langle\mathbf{P}(z^t)\sigma, X\rangle + \langle\theta, X^{-1}\rangle}\right) E\left(e^{\langle\sigma, Y\rangle}\right) \\ &= \frac{K_{-p}(\mathbf{P}(z^t)(a+\sigma), b+\theta)}{K_{-p}(\mathbf{P}(z^t)a, b)} \frac{|a|^q}{|a+\sigma|^q} \\ &= \frac{K_p(b+\theta, \mathbf{P}(z^t)(a+\sigma))}{K_p(b, \mathbf{P}(z^t)a)} \frac{|a|^q}{|a+\sigma|^q}. \end{aligned}$$

Consider now random matrices U_1, V_1 such that

$$(U_1, V_1) \sim GIG_s(-q, \mathbf{P}(z)b, a) \otimes W_r(p, b).$$

To complete the proof we need only show that $M_{(U,V)} = M_{(U_1,V_1)}$. By an argument similar to that followed above we see that

$$M_{(U_1,V_1)}(\theta, \sigma) = \frac{K_q(a+\sigma, \mathbf{P}(z)(b+\theta))}{K_q(a, \mathbf{P}(z)b)} \frac{|b|^p}{|b+\theta|^p}.$$

Now, for the case $s \leqslant r$, the result follows by Proposition 3.1.

Let us now consider the case $r < s$. Take two independent random matrices $X_1 \sim GIG_s(-q, \mathbf{P}(z)b, a)$ and $Y_1 \sim W_r(p, b)$, $X_1 \in \mathcal{V}_s^+$, $Y_1 \in \bar{\mathcal{V}}_r^+$. Note that, since $r < s$, $\mathbf{P}(z^t)X_1 + Y_1$ is positive definite. According to what we just showed,

$$(X, Y) \stackrel{d}{=} \left((\mathbf{P}(z^t)X_1 + Y_1)^{-1}, X_1^{-1} - \mathbf{P}(z)(\mathbf{P}(z^t)X_1 + Y_1)^{-1})\right). \tag{3.5}$$

Then,

$$\mathbf{P}(z)X + Y \stackrel{d}{=} X_1^{-1}$$

and therefore it is positive definite a.s. Thus U and V are well defined. Moreover, using (3.5), we have

$$(U, V) = \left((\mathbf{P}(z)X + Y)^{-1}, X^{-1} - \mathbf{P}(z^t)(\mathbf{P}(z)X + Y)^{-1}\right) \stackrel{d}{=} (X_1, Y_1). \qquad \square$$

3.2. Second proof: using the structure of the Wishart

Our basic tool here is the following proposition giving certain conditional independences for the Wishart distribution.

Proposition 3.2. *Let K be an $(r+s)\times(r+s)$ Wishart random matrix, $s\leqslant r$, with parameters $Q\in\Lambda_{r+s}$, $Q>\frac{r-1}{2}$, and $c\in\mathcal{V}^+_{r+s}$. We partition K and c in blocks according to the dimensions r and s as*

$$K=\begin{pmatrix} K_1 & K_{12}\\ K_{21} & K_2\end{pmatrix},\quad c=\begin{pmatrix} c_1 & c_{12}\\ c_{21} & c_2\end{pmatrix},$$

assuming that $c_1\in\mathcal{V}^+_r$ and $c_2\in\mathcal{V}^+_s$.

Then K_1 is of full rank and the conditional distribution of $(K_1, K_{2\cdot1})$ given K_{12} is a product of GIG and Wishart

$$(K_1, K_{2\cdot1})|K_{12}\sim GIG_r\left(Q-\frac{s}{2}, c_1, K_{12}c_2K_{21}\right)\otimes W_s\left(Q-\frac{r}{2}, c_2\right). \tag{3.6}$$

Dually, K_2 is of full rank and the conditional distribution of $(K_2, K_{1\cdot2})$ given K_{12} is GIG and Wishart

$$(K_2, K_{1\cdot2})|K_{12}\sim GIG_s\left(Q-\frac{r}{2}, c_2, K_{21}c_1K_{12}\right)\otimes W_r\left(Q-\frac{s}{2}, c_1\right). \tag{3.7}$$

Proposition 3.2 highlights the role of the matrix *GIG* distribution and the connection between the MY property and the Wishart, a theme we will study in detail in Section 5.

The proof of Proposition 3.2 can be derived from the independence of $K_{2\cdot1}$ and (K_1, K_{12}) and an extension of a result by Butler [2] giving the conditional distribution of K_1 given K_{12}. The independence of $K_{2\cdot1}$ and (K_1, K_{12}) is a well-known result for non-singular K. A proof using densities can be found for example in [16], and though it is given only for $Q=l/2, l\in\mathbf{N}, l\geqslant r+s$, it can immediately be extended for any $Q>(r+s-1)/2$ (see [12] for such a proof in a more general framework). Another proof for $Q=l/2$, $l\in\mathbf{N}$, $l\geqslant r+s$, using the normal representation of the Wishart, can be found in [4]. The assumption $l\geqslant r+s$ is redundant in that proof and the result is therefore valid for $Q=l/2$, $l\in\mathbf{N}$, $l\geqslant r$. Further details are given in the Appendix A.

Before we give the second proof of Theorem 3.1 let us make some remarks relevant to the remainder of the paper.

Remark 3.1. From Proposition 3.2, using property (2.2), we immediately get analogues of (3.6) and (3.7) respectively:

$$(K_1^{-1}, K_{2\cdot1})|K_{12}\sim GIG_r\left(-Q+\frac{s}{2}, K_{12}c_2K_{21}, c_1\right)\otimes W_s\left(Q-\frac{r}{2}, c_2\right), \tag{3.8}$$

$$(K_2^{-1}, K_{1\cdot2})|K_{12}\sim GIG_s\left(-Q+\frac{r}{2}, K_{21}c_1K_{12}, c_2\right)\otimes W_r\left(Q-\frac{s}{2}, c_1\right). \tag{3.9}$$

Remark 3.2. Observe that the *GIG* in (3.8) is well defined since $c_1\in\mathcal{V}^+_r$, $K_{12}c_2K_{21}\in\bar{\mathcal{V}}^+_r$ is of rank s and $Q-\frac{s}{2}>\frac{r-s-1}{2}$ due to our assumption $Q>\frac{r-1}{2}$. The *GIG* in (3.9) is also well defined since both $K_{21}c_1K_{12}$ and c_2 are in $\mathcal{V}^+_s$.

Remark 3.3. We note that the shape parameters of the conditional distribution of $(K_1, K_{2\cdot 1})$ and $(K_2, K_{1\cdot 2})$ given K_{12} depend on c only through blocks c_1 and c_2.

Proof of Theorem 3.1. Let us first assume that $s \leqslant r$. In this case, $\mathbf{P}(z)X + Y$ is of full rank and U and V are well defined.

Given (X, Y) as in (3.1), by Proposition 3.2 and Remark 3.1, there exists a random matrix $K \sim W_{r+s}(Q, c)$ with

$$Q = \frac{r}{2} + q = \frac{s}{2} + p,$$
$$c = \begin{pmatrix} c_1 & c_{12} \\ c_{21} & c_2 \end{pmatrix} = \begin{pmatrix} b & c_{12} \\ c_{21} & a \end{pmatrix} \in \mathcal{V}^+_{r+s}, \quad c_1 \in \mathcal{V}^+_r, \quad c_2 \in \mathcal{V}^+_s,$$

such that the conditional distribution of $(K_1^{-1}, K_{2\cdot 1})$ given $K_{12} = z^t$ is the same as the distribution of (X, Y). We note that $c_{12} = c_{21}^t$ can be chosen arbitrarily as long as $c \in \mathcal{V}^+_{r+s}$.

In order to find the distribution of (U, V) let us consider its Laplace transform. From (3.8) of Remark 3.1,

$$\begin{aligned} E\left(e^{\langle \alpha, U\rangle + \langle \beta, V\rangle}\right) &= E\left(e^{\langle \alpha, (\mathbf{P}(z)X+Y)^{-1}\rangle + \langle \beta, X^{-1} - \mathbf{P}(z^t)(\mathbf{P}(z)X+Y)^{-1}\rangle}\right) \\ &= E\left(e^{\langle \alpha, (\mathbf{P}(z)K_1^{-1}+K_{2\cdot 1})^{-1}\rangle + \langle \beta, K_1 - \mathbf{P}(z^t)(\mathbf{P}(z)K_1^{-1}+K_{2\cdot 1})^{-1}\rangle} | K_{21} = z\right) \\ &= E\left(e^{\langle \alpha, K_2^{-1}\rangle + \langle \beta, K_{1\cdot 2}\rangle} | K_{21} = z\right). \end{aligned}$$

And now from (3.9) we can conclude that (U, V) has the required distribution.

The case $r < s$ can be treated exactly as in the first proof. $\square$

4. The characterization of the *GIG* and Wishart distribution

In this section, we are going to prove a characterization of the *GIG* and Wishart distribution, that is the converse to the MY property for X and Y matrix variates of different dimensions. For $r = s$ and $z = I_r$ such a characterization has been proved in [11] under the assumption that the densities are strictly positive and twice continuously differentiable. The same characterization was given in [17] under the weaker smoothness assumptions of strict positivity and differentiability of the densities. These weaker conditions will be our smoothness assumptions here.

Theorem 4.1. *Let X and Y be two independent random matrices taking their values in $\mathcal{V}^+_r$ and $\mathcal{V}^+_s$ respectively. Assume that X and Y have strictly positive differentiable densities with respect to the Lebesgue measure. Let*

$$\psi_z(x, y) = \left((\mathbf{P}(z)x + y)^{-1}, x^{-1} - \mathbf{P}(z^t)(\mathbf{P}(z)x + y)^{-1}\right), \tag{4.1}$$

where z is a given constant $s \times r$ matrix of full rank. Let $(U, V) = \psi_z(X, Y)$.

If U and V are independent, then there exists $(a, b) \in \mathcal{V}_s^+ \times \mathcal{V}_r^+$ *and scalars p and q satisfying* $p - q = \frac{r-s}{2}$, $p > \frac{r-1}{2}$, *such that X and Y are independent GIG and Wishart with*

$$(X, Y) \sim GIG_r(-p, \mathbf{P}(z^t)a, b) \otimes W_s(q, a). \tag{4.2}$$

It also follows immediately that U and V are independent GIG and Wishart with

$$(U, V) \sim GIG_s(-q, \mathbf{P}(z)b, a) \otimes W_r(p, b) \tag{4.3}$$

Remark 4.1. Observe that if $r > s$ the parameter $\mathbf{P}(z^t)a$ of the *GIG* distribution of X is singular (semi-positive definite), but the distribution is well defined since $p > \frac{r-1}{2}$. Dually, if $r < s$ the parameter $\mathbf{P}(z)b$ of the *GIG* distribution of U is singular (semi-positive definite), but again the distribution is well defined since $q > \frac{s-1}{2}$.

Proof. Let us first note that without loss of generality we can assume that $s \geqslant r$. Indeed, as we have seen it in the proof of Lemma 2.2, $\psi_z^{-1} = \psi_{z^t}$. Therefore the couples (U, V) and $(X, Y) = \psi_{z^t}(U, V)$ play symmetric roles and we can therefore arbitrarily choose to have $s \geqslant r$.

We then have that $\mathbf{P}(z^t)(\mathcal{V}_s^+) = \mathcal{V}_r^+$. Indeed, clearly, $\mathbf{P}(z^t)(\mathcal{V}_s^+) \subset \mathcal{V}_r^+$. To prove the reverse inclusion it is sufficient to show that for any given $v \in \mathcal{V}_r^+$ there exists $u \in \mathcal{V}_s^+$ such that $\mathbf{P}(z^t)u = v$. We write $z^t = [z_1^t, z_2^t]$, where without loss of generality we can assume that z_1 is an $r \times r$ invertible matrix, while $z_2 \in \mathcal{M}_{s-r,r}$. Since $\mathcal{V}_r^+$ is an open set there exists $\varepsilon > 0$ such that $v - \varepsilon z_2^t z_2 \in \mathcal{V}_r^+$. Define the $r \times r$ matrix u_1 to be

$$u_1 = (z_1^t)^{-1}(v - \varepsilon z_2^t z_2) z_1^{-1}.$$

Then we see that

$$u = \begin{pmatrix} u_1 & 0 \\ 0 & \varepsilon I_{s-r} \end{pmatrix} \in \mathcal{V}_s^+$$

with $\mathbf{P}(z^t)u = v$ and therefore $\mathbf{P}(z^t)(\mathcal{V}_s^+) = \mathcal{V}_r^+$. This fact will be used twice in the proof below.

Let us now prove the theorem. From Lemma 2.2, we know that the Jacobian of the transformation $\psi_z^{-1}(u, v)$ is

$$J = |u|^{-(s+1)} |\mathbf{P}(z^t)u + v|^{-(r+1)}.$$

Therefore the relationship between the densities of X, Y, U and V is, for any $(u, v) \in V_s^+ \times V_r^+$,

$$f_U(u) f_V(v) = |u|^{-(s+1)} |\mathbf{P}(z^t)u + v|^{-(r+1)} f_X((\mathbf{P}(z^t)u + v)^{-1}) f_Y(y(u, v)), \tag{4.4}$$

where $y(u, v) = u^{-1} - \mathbf{P}(z)(\mathbf{P}(z^t)u + v)^{-1}$. Writing

$$\phi_1(u) = \log(f_U(u)) + (s+1)\log|u| = \tilde{\phi}_1(u^{-1}) \quad \phi_2(v) = \log f_V(v)$$
$$\phi_3(x) = \log(f_X(x)) + (r+1)\log|x| = \tilde{\phi}_3(x^{-1}) \quad \phi_4(y) = \log f_Y(y)$$

and taking logarithms of (4.4), we obtain

$$\phi_1(u) + \phi_2(v) = \tilde{\phi}_3(\mathbf{P}(z^t)u + v) + \phi_4(y(u, v)). \tag{4.5}$$

Let us now differentiate (4.5) with respect to u. We obtain

$$\begin{aligned} \phi_1'(u) &= \mathbf{P}(z)\left[\tilde{\phi}_3'(\mathbf{P}(z^t)u + v)\right] \\ &\quad + \left[\mathbf{P}(z)\mathbf{P}((\mathbf{P}(z^t)u + v)^{-1})\mathbf{P}(z^t) - \mathbf{P}(u^{-1})\right]\phi_4'(y(u, v)) \\ &= \mathbf{P}(z)\left[\tilde{\phi}_3'(\mathbf{P}(z^t)u + v)\right] \\ &\quad + \left[\mathbf{P}\left(\mathbf{P}(z)(\mathbf{P}(z^t)u + v)^{-1}\right) - \mathbf{P}(u^{-1})\right]\phi_4'(y(u, v)). \end{aligned} \tag{4.6}$$

Differentiating (4.5) with respect to v, we obtain

$$\phi_2'(v) = \tilde{\phi}_3'(\mathbf{P}(z^t)u + v) + \mathbf{P}((\mathbf{P}(z^t)u + v)^{-1})\mathbf{P}(z^t)\phi_4'(y). \tag{4.7}$$

Taking $\mathbf{P}(z)$ of (4.7) and subtracting it from (4.6), gives us

$$\phi_1'(u) - \mathbf{P}(z)\phi_2'(v) = -\mathbf{P}(u^{-1})\phi_4'(y),$$

which yields immediately

$$\phi_4'(y(u, v)) = \mathbf{P}(u)\mathbf{P}(z)\phi_2'(v) - \mathbf{P}(u)\phi_1'(u). \tag{4.8}$$

We can rewrite (4.5) in terms of x and y as follows

$$\tilde{\phi}_1(\mathbf{P}(z)x + y) + \phi_2(v(x, y)) = \phi_3(x) + \phi_4(y).$$

Following steps parallel to the previous ones we differentiate with respect to x and y successively. We obtain

$$\phi_2'(v(x, y)) = \mathbf{P}(x)\mathbf{P}(z^t)\phi_4'(y) - \mathbf{P}(x)\phi_3'(x). \tag{4.9}$$

We find it convenient to introduce the notation:

$$\psi_1(u) = \mathbf{P}(u)\phi_1'(u), \quad \psi_2(v) = \phi_2'(v), \quad \psi_3(x) = \mathbf{P}(x)\phi_3'(x), \quad \psi_4(y) = \phi_4'(y)$$

so that (4.8) and (4.9) can be rewritten as

$$\psi_4(y) = \mathbf{P}(u)\mathbf{P}(z)\psi_2(v) - \psi_1(u), \tag{4.10}$$

$$\psi_2(v) = \mathbf{P}(x)\mathbf{P}(z^t)\psi_4(y) - \psi_3(x). \tag{4.11}$$

Multiplying (4.10) by $\mathbf{P}(x)\mathbf{P}(z^t)$ and subtracting (4.11) from it, we obtain

$$\begin{aligned} \psi_3(x) &= \left(\mathbf{P}(x)\mathbf{P}(z^t)\mathbf{P}(u)\mathbf{P}(z) - \mathrm{Id}_r\right)\psi_2(v) - \mathbf{P}(x)\mathbf{P}(z^t)\psi_1(u) \\ &= \left(\mathbf{P}(x)\mathbf{P}(\mathbf{P}(z^t)u) - \mathrm{Id}_r\right)\psi_2(v) - \mathbf{P}(x)\mathbf{P}(z^t)\psi_1(u), \end{aligned}$$

which we rewrite as

$$\begin{aligned} \mathbf{P}(x^{-1})\psi_3(x) &= \left(\mathbf{P}(\mathbf{P}(z^t)u) - \mathbf{P}(x^{-1})\right)\psi_2(v) - \mathbf{P}(z^t)\psi_1(u) \\ &= \left(\mathbf{P}(\mathbf{P}(z^t)u) - \mathbf{P}(\mathbf{P}(z^t)u + v)\right)\psi_2(v) - \mathbf{P}(z^t)\psi_1(u). \end{aligned} \tag{4.12}$$

Since for any fixed v, $x = (\mathbf{P}(z^t)u + v)^{-1}$ is a function of $\tilde{u} = \mathbf{P}(z^t)u$, it follows that $\mathbf{P}(z^t)\psi_1(u)$ is a function A of $\tilde{u}$ and therefore (4) is of the form

$$C(\tilde{u} + v) = A(\tilde{u}) + [\mathbf{P}(\tilde{u} + v) - \mathbf{P}(\tilde{u})]\psi_2(v) \tag{4.13}$$

holding for all $\tilde{u},\ v \in \mathcal{V}_r^+$, where $C(x) = -\mathbf{P}(x)\psi_3(x^{-1})$.

We now need the following theorem given in [17].

Theorem 4.2. *Let $\mathcal{V}$ be the space of symmetric matrices of a given dimension d and $\mathcal{V}^+$ be its cone of positive definite matrices. Also let $A, B : \mathcal{V}^+ \to \mathcal{V}$ be some functions and $C : \mathcal{V}^+ \times \mathcal{V}^+ \to \mathcal{V}$ be a symmetric function, that is, $C(u, v) = C(v, u)$ for any $u, v \in \mathcal{V}^+$. If*

$$A(u) + [\mathbf{P}(u + v) - \mathbf{P}(u)]B(v) = C(u, v)$$

holds for any $u, v \in \mathcal{V}^+$, then there exist $a, b \in \mathcal{V}$ and $\lambda \in \mathbf{R}$ such that for any $u, v \in \mathcal{V}^+$

$$A(u) = a - \lambda u + \mathbf{P}(u)b\ , \quad B(v) = b + \lambda v^{-1},$$

$$C(u, v) = a + \lambda(u + v) + \mathbf{P}(u + v)b.$$

We apply this theorem with $d = r$ and $\mathcal{V}^+ = \mathcal{V}_r^+$ to Eq. (4.13) and it follows immediately that there exist $\tilde{a}, b \in \mathcal{V}_r$ and a scalar λ such that

$$\phi_2'(v) = -b + \lambda v^{-1}\ , \quad \phi_3'(x) = -\tilde{a} - \lambda x^{-1} + \mathbf{P}(x^{-1})b.$$

Taking antiderivatives yields the densities for V and X as follows:

$$f_V(v) = c_V|v|^{\lambda}\exp(-\langle b, v\rangle),$$

$$f_X(x) = c_X|x|^{-(\lambda+r+1)}\exp(-\langle \tilde{a}, x\rangle - \langle b, x^{-1}\rangle),$$

where c_X and c_V are appropriate constants. Since f_V and f_X are densities of probability measures it follows that $\tilde{a}$ and b belong to $\mathcal{V}_r^+$ and there exists $p > \frac{r-1}{2}$ such that $\lambda = p - \frac{r+1}{2}$ so that $\lambda + r + 1 = p + \frac{r+1}{2}$. As noted before any $\tilde{a} \in \mathcal{V}_r^+$ can be written as $\tilde{a} = \mathbf{P}(z^t)a$ for some $a \in \mathcal{V}_s^+$. Therefore we have proved that

$$X \sim GIG_r(-p, \mathbf{P}(z^t)a, b) \quad \text{and} \quad V \sim W_r(p, b).$$

Writing the densities of X and V in (4.4) gives

$$\begin{aligned} &f_U(u)c_V|v|^{p-\frac{r+1}{2}}\exp -\langle b, v\rangle \\ &\quad = |u|^{-(s+1)}|x|^{(r+1)}c_X|x|^{-p-\frac{r+1}{2}}\exp\left(-\langle \mathbf{P}(z^t)a, x\rangle - \langle b, x^{-1}\rangle\right) f_Y(y). \end{aligned}$$

Using the fact that $|x||y|^{-1} = |u||v|^{-1}$ and defining the scalar q such that

$$p - q = \frac{r - s}{2},$$

the preceding equation gives us

$$\begin{aligned} f_U(u) &= C|u|^{-(s+1)}|x|^{-p+\frac{r+1}{2}}|v|^{-p+\frac{r+1}{2}} \exp\left(-\langle \mathbf{P}(z^t)a, x\rangle - \langle b, x^{-1} - v\rangle\right) f_Y(y) \\ &= C|u|^{-(s+1)}(|u||y|)^{-p+\frac{r+1}{2}} \exp\left(-\langle \mathbf{P}(z^t)a, x\rangle - \langle b, \mathbf{P}(z^t)u\rangle\right) f_Y(y) \\ &= C|u|^{-p+\frac{r-s}{2}-\frac{s+1}{2}}|y|^{-p+\frac{r+1}{2}} \exp\left(-\langle \mathbf{P}(z)b, u\rangle - \langle a, \mathbf{P}(z)x\rangle\right) f_Y(y) \\ &= C|u|^{-q-\frac{s+1}{2}}|y|^{-p+\frac{r+1}{2}} \exp\left(-\langle \mathbf{P}(z)b, u\rangle - \langle a, u^{-1} - y\rangle\right) f_Y(y) \\ &= C|u|^{-q-\frac{s+1}{2}} \exp\left(-\langle \mathbf{P}(z)b, u\rangle - \langle a, u^{-1}\rangle\right) |y|^{-p+\frac{r+1}{2}} \exp\left(\langle a, y\rangle\right) f_Y(y), \end{aligned}$$

where C is an appropriate constant. By the principle of separation of variables and since $-p + \frac{r+1}{2} = -q + \frac{s+1}{2}$, it follows that the density f_Y of Y is that of the $W_s(q, a)$ and the density f_U of U is that of the $GIG_s(-q, \mathbf{P}(z)b, a)$. $\square$

5. A characterization of the Wishart distribution through independence of blocks

The link between the conditional structure of the Wishart distribution and the MY property was first used in the case of 2×2 matrices in [10]. This link carries on here for the extended MY property and Wishart matrices of dimension $n \times n$, $n \geqslant 3$. As was observed in [6] it is not possible to characterize the Wishart distribution for $n = 2$ and indeed in [10] the MY property was linked to the quasi-Wishart distribution only. Geiger and Heckerman [6] gave a characterization of the Wishart for $n \geqslant 3$ assuming independences for all possible block partitionings of the random matrix considered. In our characterization, given in Theorem 5.1 below, the importance of the fact that n must be greater than or equal to 3 in order to obtain a characterization of the Wishart is indicated by the fact that we need independences for only three different block partitionings of the random matrix.

For an $n \times n$ matrix k and a given $i \in \{1, \ldots, n\}$, define the partitioning into blocks $(k_{i,(1)}, k_{i,(12)}, k_{i,(2)})$, where

$$k_{i,(1)} = [k_{ii}]\,, \quad k_{i,(12)} = [k_{ij}]_{j \neq i} = k^t_{i,(21)}\,, \quad k_{i,(2)} = [k_{lm}]_{l \neq i \text{ and } m \neq i}. \tag{5.1}$$

The dimensions of the blocks are 1×1, $1 \times (n-1)$ and $(n-1) \times (n-1)$, respectively. Our assumption, in this section, is that k belongs to $\mathcal{V}_n^+$ and therefore ${k_{i,(1)}}^{-1}$ and ${k_{i,(2)}}^{-1}$ exist. We can then use the following notation

$$\begin{aligned} k_{i,(2)\cdot(1)} &= k_{i,(2)} - k_{i,(21)}{k_{i,(1)}}^{-1}k_{i,(12)}, \\ k_{i,(1)\cdot(2)} &= k_{i,(1)} - k_{i,(12)}{k_{i,(2)}}^{-1}k_{i,(21)}. \end{aligned}$$

Theorem 5.1. *Let K be an $n \times n$ random matrix taking its values in $\mathcal{V}_n^+$ having a strictly positive differentiable density. Then K is Wishart distributed if and only if*

$$(K_{i,(1)}, K_{i,(12)}) \quad \textit{and} \quad K_{i,(2)\cdot(1)} \quad \textit{are independent} \tag{5.2}$$

and also

$$(K_{i,(2)}, K_{i,(12)}) \quad \textit{and} \quad K_{i,(1)\cdot(2)} \quad \textit{are independent} \tag{5.3}$$

for three distinct values of $i \in \{1, \ldots, n\}$.

Proof. If K is non-singular Wishart then it is well known that these independences are satisfied for any block partitioning of K.

Let us now prove the converse. For any i such that (5.2) and (5.3) hold, let us consider the random variates (X, Y) having the same distribution as the conditional distribution of $({K_{i,(1)}}^{-1}, K_{i,(2)\cdot(1)})$ given $K_{i,(12)} = z^t$. Then $(U, V) = \psi_z(X, Y)$ follows the same law as the conditional law of $({K_{i,(2)}}^{-1}, K_{i,(1)\cdot(2)})$ given $K_{i,(12)} = z^t$ for any given z of appropriate dimension. This can be seen for instance through a Laplace transform argument as in the second proof of Theorem 3.1. By (5.2), X and Y are independent and by (5.3), U and V are independent. Then by Theorem 4.1, there exist scalars p_i and q_i and constant matrices $a_i \in \mathcal{V}_s^+$, $b_i \in \mathcal{V}_r^+$ satisfying the conditions given in Theorem 4.1, such that, conditionally on $K_{i,(12)}$, we have

$$\begin{aligned}&\left[\left({K_{i,(1)}}^{-1}, K_{i,(2)\cdot(1)}\right) \mid K_{i,(12)} = z^t\right]\\ &\quad\sim\ GIG_1(-p_i, P(z^t)a_i, b_i) \otimes W_{n-1}(q_i, a_i)\end{aligned} \tag{5.4}$$

and

$$\begin{aligned}&\left[\left({K_{i,(2)}}^{-1}, K_{i,(1)\cdot(2)}\right) \mid K_{i,(12)} = z^t\right]\\ &\quad\sim\ GIG_{n-1}(-q_i, P(z)b_i, a_i) \otimes W_1(p_i, b_i).\end{aligned} \tag{5.5}$$

Since the Wishart distributions $W_{n-1}(q_i, a_i)$ and $W_1(p_i, b_i)$ are the conditional distributions of $K_{i,(2)\cdot(1)}$ given $K_{i,(12)} = z^t$ and $K_{i,(1)\cdot(2)}$ given $K_{i,(12)} = z^t$, respectively, the parameters p_i, q_i, a_i and b_i may be dependent upon z. However, $K_{i,(2)\cdot(1)}$ and $K_{i,(12)}$ are independent and so are $K_{i,(1)\cdot(2)}$ and $K_{i,(12)}$ and therefore these parameters are constants.

Without loss of generality we can assume that the three values of i for which the independences hold are $i = 1, 2, 3$.

To go further in our proof we need to set some notation:

$$\begin{aligned}
K_{-1} &= (K_{2,3},\ K_{2,4}, \ldots,\ K_{2,n},\ K_{3,4}, \ldots,\ K_{3,n}, \ldots,\ K_{n-1,n}),\\
K_{-2} &= (K_{1,3},\ K_{1,4}, \ldots,\ K_{1,n},\ K_{3,4}, \ldots,\ K_{3,n}, \ldots,\ K_{n-1,n}),\\
K_{-3} &= (K_{1,2},\ K_{1,4}, \ldots,\ K_{1,n},\ K_{2,4}, \ldots,\ K_{2,n},\ K_{4,5}, \ldots\ K_{4,n}, \ldots,\ K_{n-1,n}).
\end{aligned}$$

Following the definition of $k_{i,(12)}$ in (5.1) we make the following identifications

$$\begin{aligned}
K_{1,(12)} &= (K_{1,2},\ K_{1,3}, \ldots,\ K_{1,n}),\\
K_{2,(12)} &= (K_{1,2},\ K_{2,3}, \ldots,\ K_{2,n}),\\
K_{3,(12)} &= (K_{1,3},\ K_{2,3},\ K_{3,4}, \ldots,\ K_{3,n}).
\end{aligned}$$

At this point let us note that, for each $i = 1, 2, 3$ the pair $(K_{i,(12)}, K_{-i})$ consists of all the off-diagonal elements in the upper triangular part of K. We denote this set of elements as K_{-d}. From (5.5) we know that the conditional distribution of $K_{i,(2)}^{-1}$ given $K_{i,(12)}$ is $GIG_{n-1}(q_i, P(K_{i,(21)})b_i, a_i)$. Thus by (2.2), the conditional distribution

$$\left(K_{i,(2)} \mid K_{i,(12)}\right) \sim GIG_{n-1}(-q_i, a_i, P(K_{i,(21)})b_i)$$

and therefore we know the marginal conditional distribution of K_{-i} given $K_{i,(12)}$. Their densities can be written as

$$f_{-i|i,(12)}(k_{-i}|k_{i,(12)}) = c_i(k_{-d}) = \frac{f(k_{-d})}{f_i(k_{i,(12)})}, \tag{5.6}$$

where c_i is known, $i = 1, 2, 3$. Therefore

$$c_1(k_{-d})f_1(k_{1,(12)}) = c_2(k_{-d})f_2(k_{2,(12)}) = c_3(k_{-d})f_3(k_{3,(12)}). \tag{5.7}$$

Since $k_{-d} = (k_{1,(12)}, k_{-1})$, then setting $k_{1,(12)} = (0, \ldots, 0) = \mathbf{0}$ in the first equality of (5.7), we get

$$c_1(\mathbf{0}, k_{-1})f_1(\mathbf{0}) = c_2(\mathbf{0}, k_{-1})f_2(0, k_{2,3}, \ldots, k_{2,n}). \tag{5.8}$$

Setting $k_{1,2} = 0$ in the second equality of (5.7) we obtain

$$\begin{aligned} &c_2(k_{1,2} = 0, k_{1,3}, \ldots, k_{1,n}, k_{2,3}, \ldots, k_{2,n}, \ldots, k_{n-1,n})f_2(0, k_{2,3}, \ldots, k_{2,n}) \\ &\quad = c_3(k_{1,2} = 0, k_{1,3}, \ldots, k_{1,n}, k_{2,3}, \ldots, k_{2,n}, \ldots, k_{n-1,n})f_3(k_{3,(12)}). \end{aligned} \tag{5.9}$$

Since it is assumed that the density is non-zero then

$$c_2(\mathbf{0}, k_{-1}) \neq 0 \text{ and } c_3(0, k_{1,3}, \ldots, k_{1,n}, k_{2,3}, \ldots, k_{2,n}, \ldots, k_{n-1,n}) \neq 0.$$

Then, combining (5.8) and (5.9) we see that the density $f_3(k_{3,(12)})$ is uniquely determined by the functions c_1, c_2 and c_3.

From (5.5) with $i = 3$ it now follows that the joint distribution of

$$\left(K_{3,(2)}, K_{3,(1)\cdot(2)}, K_{3,(12)}\right)$$

is uniquely identified by the set of parameters $\{a_i, b_i, p_i, q_i, i = 1, 2, 3\}$. We have therefore uniquely identified the distribution of K.

Since the Wishart distribution satisfies the independence conditions in our assumptions the distribution of K must be Wishart. □

Appendix A.

A.1. Proof of Proposition 3.1

For the proof of Proposition 3.1 we need the following lemma.

Lemma A.1. *Let $x \in \mathcal{V}_r^+$ and $y \in \bar{\mathcal{V}}_r^+$ with $rank(y) = s \in \{0, 1, \ldots, r-1\}$. Denote by y_1, y_{12}, y_2 the block decomposition of the matrix y with respective dimensions of the blocks: $(r-s) \times (r-s)$, $(r-s) \times s$ and $s \times s$, and assume that $y_2 \in \mathcal{V}_s^+$. Let $p > \frac{r-s-1}{2}$ and $q = p - \frac{r-s}{2}$.*

Then

$$K_p(x,y) = \Gamma_{r-s}(p)\pi^{\frac{r(r-s)}{2}}|x|^{-p}\left|(\mathbf{P}(\rho_y)x)_2\right|^q K_q\left((\mathbf{P}(\rho_y)x)_2, y_2\right), \tag{A.1}$$

where

$$\rho_y = \begin{pmatrix} I_{r-s} & 0 \\ y_2^{-1}y_{21} & I_s \end{pmatrix}$$

and $(d)_2$ *denotes the diagonal block* d_{22} *of the matrix* d.

Proof. We first observe that

$$y = \rho_y^t \tilde{y}_2 \rho_y, \qquad \text{with } \tilde{y}_2 = \begin{pmatrix} 0 & 0 \\ 0 & y_2 \end{pmatrix}.$$

Thus

$$K_p(x,y) = \int_{\mathcal{V}_r^+} |u|^{p-\frac{r+1}{2}} e^{-\langle x,u\rangle - \langle \tilde{y}_2, \rho_y u^{-1}\rho_y^t\rangle}\, du.$$

In the above integral, we make the change of variable $u \mapsto v = (\rho_y^t)^{-1}u\rho_y^{-1}$. The Jacobian is clearly 1 and, since $\langle \tilde{y}_2, v^{-1}\rangle = \langle y_2, (v^{-1})_2\rangle = \langle y_2, v_{2\cdot 1}^{-1}\rangle$, we obtain

$$K_p(x,y) = \int_{\mathcal{V}_r^+} |v|^{p-\frac{r+1}{2}} e^{-\langle \tilde{x},v\rangle - \langle y_2, v_{2\cdot 1}^{-1}\rangle}\, dv,$$

where $\tilde{x} = \rho_y x \rho_y^t$.

Now using the identity

$$\langle \tilde{x}, v\rangle = \langle \tilde{x}_2, v_{2\cdot 1}\rangle + \langle \tilde{x}_{1\cdot 2}, v_1\rangle + \langle \tilde{x}_{12}\tilde{x}_2^{-1} + v_1^{-1}v_{12}, \tilde{x}_2(\tilde{x}_2^{-1}\tilde{x}_{21} + v_{21}v_1^{-1})v_1\rangle \tag{A.2}$$

and making the change of variables

$$v = (v_1, v_{12}, v_2) \mapsto w = (w_1 = v_1, w_{12} = v_1^{-1}v_{12}, w_2 = v_{2\cdot 1}),$$

the Jacobian of which is $J = |w_1|^s$, after a routine calculation we obtain (A.1). □

Proof of Proposition 3.1. We first consider the case of $p > \frac{r-1}{2}$, that is $q > \frac{s-1}{2}$. In this case, by (2.3), we have

$$\begin{aligned} \Gamma_s(q)|a|^{-q}K_p(b, \mathbf{P}(z^t)a) &= \Gamma_s(q)|a|^{-q}K_{-p}(\mathbf{P}(z^t)a, b) \\ &= \int_{\mathcal{V}_s^+\times\mathcal{V}_r^+} |y|^{q-\frac{s+1}{2}}|x|^{p-\frac{r+1}{2}} e^{-\langle a, y+\mathbf{P}(z)x\rangle - \langle b, x^{-1}\rangle}\, dy\, dx. \end{aligned}$$

Making the change of variables

$$u = (\mathbf{P}(z)x + y)^{-1} \in \mathcal{V}_s^+, \quad v = x^{-1} - \mathbf{P}(z^t)[\mathbf{P}(z)x+y]^{-1} \in \mathcal{V}_r^+,$$

which, by Lemma 2.2, is a bijection with Jacobian of the inverse equal to $|u|^{-(s+1)}|x|^{r+1}$, and remembering that $|x|^{-1}|y| = |u|^{-1}|v|$ we obtain

$$\Gamma_s(q)|a|^{-q}K_p(b, \mathbf{P}(z^t)a) = K_{-q}(\mathbf{P}(z)b, a)\Gamma_r(p)|b|^{-p}.$$

The result now follows by (3.4).

Let us now consider the case where $p \in \left(\frac{r-s-1}{2}, \frac{r-1}{2}\right]$. We are going to use Lemma A.1. We therefore need to identify $y = \mathbf{P}(z^t)a$ and the corresponding ρ_y. Without loss of generality we can assume that z can be decomposed into blocks $z = [z_1, z_2]$ of dimensions $s \times (r-s)$ and $s \times s$ respectively where z_2 is of rank s. Since

$$y = \mathbf{P}(z^t)a = \begin{pmatrix} z_1^t a z_1 & z_1^t a z_2 \\ z_2^t a z_1 & z_2^t a z_2 \end{pmatrix}$$

then

$$\rho_y = \begin{pmatrix} I_{r-s} & 0 \\ z_2^{-1} a^{-1} (z_2^t)^{-1} z_2^t a z_1 & I_s \end{pmatrix} = \begin{pmatrix} I_{r-s} & 0 \\ z_2^{-1} z_1 & I_s \end{pmatrix},$$

and

$$\begin{aligned} \left(\mathbf{P}(\rho_y)b\right)_2 &= \left(\begin{pmatrix} I_{r-s} & 0 \\ z_2^{-1} z_1 & I_s \end{pmatrix}\begin{pmatrix} b_1 & b_{12} \\ b_{21} & b_2 \end{pmatrix}\begin{pmatrix} I_{r-s} & z_1^t (z_2^{-1})^t \\ 0 & I_s \end{pmatrix}\right)_2 \\ &= z_2^{-1}(\mathbf{P}(z)b)(z_2^{-1})^t. \end{aligned}$$

Thus by Lemma A.1 we obtain

$$\begin{aligned} K_p(b, \mathbf{P}(z^t)a) &= \Gamma_{r-s}(p)\pi^{\frac{(r-s)s}{2}} |b|^{-p} |z_2|^{-2q} |(\mathbf{P}(z)b)|^q \\ &\quad \times K_q\left(z_2^{-1}(\mathbf{P}(z)b)(z_2^{-1})^t, z_2^t a z_2\right). \end{aligned}$$

Since $K_q(z_2^{-1}c(z_2^{-1})^t, z_2^t a z_2) = |z_2|^{2q} K_q(c, a)$ we obtain

$$K_p(b, \mathbf{P}(z^t)a) = \frac{\Gamma_r(p)}{\Gamma_s(q)} |b|^{-p} |\mathbf{P}(z)b|^q K_q(\mathbf{P}(z)b, a).$$

Now (3.3) follows by (3.4). □

A.2. Proof of Proposition 3.2

Proof of Proposition 3.2. As mentioned in Section 3, the independence of $K_{2\cdot 1}$ and (K_1, K_{12}) when K is Wishart $W_{r+s}(Q, c)$ has been proved or can be immediately derived from existing proofs, for all values of $Q \in \Lambda_{r+s}$. The conditional independences (3.6) and (3.7) then follow by (2.13). The distributions of $K_{2\cdot 1}$ and $K_{1\cdot 2}$ which are, of course, also conditional distributions given K_{12} are also well known to be Wishart as given in (3.6) and (3.7). It only remains to show that the conditional distributions of K_1 and K_2 given K_{12} are *GIG*'s with appropriate parameters. This result for the non-singular case, that is for $Q > \frac{r+s-1}{2}$, is given in Theorem 1 of Butler [2]. The singular case was considered in his Theorem 2 which we can apply directly to derive the conditional distribution of K_2 given K_{12} as given in (3.7) since both matrix parameters are then positive definite. Theorem 2 can also be directly applied to derive the conditional distribution of K_1 given K_{12} as given in (3.8) in the particular case that $r = s$. However, for the conditional distribution of K_1 given K_{12}, in the case that $s < r$, the second matrix parameters of the *GIG* is not of full rank. Then the range of the scalar parameter for the *GIG* given in [2] does not coincide with the range given in Section 2 of the present paper.

To clarify the situation let us therefore consider the case $Q \in (\frac{r-1}{2}, \frac{r+s-1}{2}] \cap \Lambda_{r+s}$ in more detail. Following a classical approach, as done in [2], we write $N = 2Q$ and take X and Y to be two $N \times r$ and $N \times s$ matrices, respectively, such that the N rows of the matrix $N \times (r+s)$ matrix $[X, Y]$ are i.i.d. $\mathcal{N}_{r+s}(0, \Sigma)$ random vectors. Then

$$K = \begin{pmatrix} K_1 & K_{12} \\ K_{21} & K_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} X^t \\ Y^t \end{pmatrix} \begin{pmatrix} X & Y \end{pmatrix} = \begin{pmatrix} X^t X & X^t Y \\ Y^t X & Y^t Y \end{pmatrix},$$

where $\stackrel{d}{=}$ denotes identity of the distributions.

It is well known that

$$X^t Y \mid X \sim \mathcal{N}_{r,s} \left(X^t X \Sigma_1^{-1} \Sigma_{12},\ \Sigma_{2\cdot 1} \otimes X^t X \right). \tag{A.3}$$

Since the conditional distribution of $X^t Y$ given X depends on X only through $X^t X$, for any $\theta \in \mathcal{M}_{r,s}$ the Laplace transform of the conditional distribution of $X^t Y$ given $X^t X$ is

$$\mathbf{E}(e^{\langle \theta, X^t Y \rangle} | X^t X) = \mathbf{E}(\mathbf{E}(e^{\langle \theta, X^t Y \rangle} | X) | X^t X) = \mathbf{E}(\phi_\theta(X^t X) | X^t X) = \phi_\theta(X^t X), \tag{A.4}$$

where $\phi_\theta(X^t X) = \mathbf{E}(e^{\langle \theta, X^t Y \rangle} | X)$. Thus the conditional distribution of $X^t Y$ given X is the same as the conditional distribution of $X^t Y$ given $X^t X$. Now the distribution of $X^t X$ is known and we therefore have the joint distribution of K_{12}, K_1 and proceed from there as in Appendix A of Butler [2] to obtain the conditional distribution of K_1 given K_{12} as given in (3.6). $\square$

References

[1] O. Barndorff-Nielsen, P. Blæsild, J.L. Jensen, B. Jørgensen, Exponential transformation models, Proc. Roy. Soc. London A 379 (1982) 41–65.

[2] R.W. Butler, Generalized inverse Gaussian distributions and their Wishart connections, Scand. J. Statist. 25 (1998) 69–75.

[3] M. Casalis, G. Letac, Characterization of the Jorgensen set in generalized linear models, Test 3 (1994) 145–162.

[4] M.L. Eaton, Multivariate Statistics—A Vector Space Approach, Wiley, New York, 1983.

[5] D. Geiger, D. Heckerman, A characterization of the bivariate Wishart distribution, Probab. Math. Statist. 18 (1998) 119–131.

[6] D. Geiger, D. Heckerman, Parameter priors for directed acyclic graphical models and the characterization of several probability distributions, Ann. Statist. 30 (2002) 1412–1440.

[7] S.G. Gindikin, Invariant generalized functions in homogeneous domains, Functional Anal. Appl. 9 (1975) 50–52.

[8] C.S. Herz, Bessel functions of matrix argument, Ann. Math. 61 (1955) 474–523.

[9] G. Letac, Convergence de l'intégrale GIG, unpublished note, 2003, pp. 1–2.

[10] G. Letac, H. Massam, The normal quasi-Wishart distribution, in: M.A.G. Viana, D.St.P. Richards (Eds.), Algebraic Methods in Statistics and Probability, AMS Contemp. Math., 287 (2001) 231–239.

[11] G. Letac, J. Wesołowski, An independence property for the product of GIG and gamma laws, Ann. Probab. 28 (2000) 1371–1383.

[12] H. Massam, E. Neher, On transformations and determinants of Wishart variables on symmetric cones, J. Theor. Probab. 10 (1997) 867–902.

[13] H. Massam, J. Wesołowski, The Matsumoto–Yor property on trees, Bernoulli 10 (2004) 685–700.

[14] H. Matsumoto, M. Yor, An analogue of Pitman's $2M$-X theorem for exponential Wiener functionals, part II: the role of the generalized inverse Gaussian laws, Nagoya Math. J. 162 (2001) 65–86.

[15] H. Matsumoto, M. Yor, Interpretation via Brownian motion of some independence properties between GIG and gamma variables, Statist. Probab. Lett. 61 (2003) 253–259.
[16] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[17] J. Wesołowski, The Matsumoto–Yor independence property for GIG and Gamma laws, revisited, Math. Proc. Cambridge Philos. Soc. 133 (2002) 153–161.