Metrika (2008) 68:219–232
DOI 10.1007/s00184-007-0154-3

More on connections between Wishart and matrix GIG distributions

V. Seshadri · J. Wesołowski

Received: 17 September 2006 / Published online: 17 October 2007
© Springer-Verlag 2007

Abstract The paper is devoted to relations between the matrix GIG and Wishart distributions. Our basic tool in the first part is a version of the Matsumoto-Yor property for matrix variables. This approach covers the following issues: the Herz identity for the Bessel function of matrix variate argument, characterization of a class of Wishart matrices and linear transformations of the matrix GIG distribution. The Bayesian Wishart model, studied in the second part, gives an alternative definition of the matrix GIG distribution. Such a model is characterized by linearity of conditional expectations and matrix GIG conditional distribution. It is also extended to Bayesian matrix GIG models, in the framework of which an interesting independence property is proved.

Keywords Wishart matrix · GIG distribution

1 Introduction

Let $\mathcal{V}_n$ be the Euclidean space of $n \times n$ real symmetric matrices equipped with the inner product $\langle a, b\rangle = \text{trace}\,(ab)$. 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 $q \in \Lambda_n = \{0, \frac{1}{2}, \frac{2}{2}, \dots, \frac{n-1}{2}\} \cup (\frac{n-1}{2}, \infty)$. 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

V. Seshadri
Department of Mathematics, McGill University, Montreal, Canada
e-mail: vansesh@hotmail.com

J. Wesołowski (✉)
Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, Poland
e-mail: wesolo@mini.pw.edu.pl

its Laplace transform is

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

Note that here we 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 Gindikin 1975 or Casalis and Letac 1994). When $q > \frac{n-1}{2}$, that is when Y takes its values in $\mathcal{V}_n^+$, this distribution has density of the form

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

where Γ_n is the multivariate Gamma function, see Muirhead (1982). When $q \in \Lambda_n$ and $q < \frac{n-1}{2}$ the distribution is singular and is concentrated on the boundary of $\overline{\mathcal{V}}_n^+$. 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 matrix generalized inverse Gaussian distribution, $\mathrm{MGIG}_n(-p, a, b)$, if it has density of the form

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

where $\mathcal{K}_p^{(n)}$ is the matrix variate modified Bessel function of the third kind, see Herz (1955). (The superscript (n) indicating the dimension of the matrix arguments will be omitted at certain places in the sequel, when the dimension will be obvious from the context.) Letac (2003) has observed that the $\mathrm{MGIG}_n(-p, a, b)$ 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. $a \in \overline{\mathcal{V}}_n^+$ with $\mathrm{rank}(a) = m \in \{0, 1, \ldots, n-1\}$, $b \in \mathcal{V}_n^+$ and $p > \frac{n-m-1}{2}$,
3. $a \in \mathcal{V}_n^+$, $b \in \overline{\mathcal{V}}_n^+$ with $\mathrm{rank}(b) = m \in \{0, 1, \ldots, n-1\}$ and $p < -\frac{n-m-1}{2}$.

This extends the original definition of the matrix variate GIG as given in Barndorff-Nielsen et al. (1982). This definition was given in the setting of exponential transformation models, and derived MGIG as an extension of the Wishart exponential model by a special affine transformation and a special statistics—see Example 4 of that paper. In this sense the first connection between Wishart and MGIG is the moment of the birth of the MGIG.

It is immediate to see that the MGIG distribution has the following property

$$\text{if } X \sim \mathrm{MGIG}_n(-p, a, b) \quad \text{then } X^{-1} \sim \mathrm{MGIG}_n(p, b, a), \tag{2}$$

which can be rephrased in terms of Bessel functions as

$$\mathcal{K}_{-p}(a, b) = \mathcal{K}_p(b, a). \tag{3}$$

For $X \sim \mathrm{GIG}_n(p, a, b)$ and for $a - \theta$ and $b - \sigma$ satisfying one of conditions 1–3 above, we have the following obvious identity

$$E(\exp(\langle \theta, X \rangle + \langle \sigma, X^{-1} \rangle)) = \frac{\mathcal{K}_p(a-\theta, b-\sigma)}{\mathcal{K}_p(a, b)}. \tag{4}$$

Note that for $\sigma = 0$ it is the Laplace transform of the MGIG distribution.

There are several connections between the Wishart and MGIG distributions known in the literature. Some of them follow the pattern of relations between univariate gamma and generalized inverse Gaussian distribution. But even then some care has to be taken of the automatic analogy between the univariate and matrix-variate situations. Some of them, due to their nature, do not possess univariate counterparts.

Wishart $W(p, a)$ is a limiting case of $\mathrm{MGIG}(p, a, b)$ as $b \to 0$, at least for $p > \frac{n-1}{2}$. It follows just by inspecting the density and the observation that, via the Lebesgue dominated convergence theorem, for $p > \frac{n-1}{2}$ one has

$$\lim_{b \to 0} \mathcal{K}_p(a, b) = \frac{\Gamma_n(p)}{|a|^p},$$

where the convergence of $b \in \mathcal{V}_n^+$ to the zero matrix is taken, for instance, in the sense of the Euclidean norm in the space $\mathcal{V}^+$.

However it is not clear what happens if p is in the discrete part of the Gindikin set Λ_n or outside of this set.

In Bernadac (1995) a matrix variate version of the representation of the GIG random variable by a continued random fraction of independent gamma variables and related characterizations obtained by Letac and Seshadri (1983) was given. Her main result is the following

Theorem 1 *Let X, Y_1 and Y_2 be independent random matrices valued in $\mathcal{V}_n^+$, such that $Y_1 \sim W(p, a)$, $Y_2 \sim W(p, b)$ with $p > \frac{n-1}{2}$ and $a, b \in \mathcal{V}_n^+$. Then $X \sim MGIG(-p, a, b)$ iff*

$$X \stackrel{d}{=} (Y_1 + (Y_2 + X)^{-1})^{-1}.$$

It was Butler (1998) who first pointed out that MGIG is incorporated in the inner structure of Wishart matrices and if the Wishart matrix is partitioned into blocks $\begin{pmatrix} K_1 & K_{12} \\ K_{21} & K_2 \end{pmatrix}$ then the conditional distribution of K_1 given K_{12} is MGIG. This relation was elaborated on further in Massam and Wesołowski (2006), where the authors gave a rigorous proof of the following result extending Butler's observation. (The symbol $\otimes$ used in the formulation denotes the product of measures.)

Theorem 2 *Let K be an $(r+s) \times (r+s)$ Wishart random matrix, $s \leq r$, 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* $(K_1, K_{2\cdot 1})$ *given* K_{12} *is a product of MGIG and Wishart*

$$(K_1, K_{2\cdot 1})|K_{12} \sim \mathrm{MGIG}_r\left(Q - \frac{s}{2}, c_1, K_{12}c_2K_{21}\right) \otimes W_s\left(Q - \frac{r}{2}, c_2\right). \tag{5}$$

Dually, K_2 *is of full rank and the conditional distribution of* $(K_2, K_{1\cdot 2})$ *given* $K_{12}(= K_{21}^T)$ *is a product of MGIG and Wishart*

$$(K_2, K_{1\cdot 2})|K_{12} \sim \mathrm{MGIG}_s\left(Q - \frac{r}{2}, c_2, K_{21}c_1K_{12}\right) \otimes W_r\left(Q - \frac{s}{2}, c_1\right). \tag{6}$$

It was done in the context of searching for the Matsumoto-Yor (MY) property, which is another interesting connection between the MGIG and Wishart distributions. For the first time this relation was proved in Letac and Wesołowski (2000) for matrices of the same dimensions together with a related characterization (this part was then extended in Wesołowski 2002). In full generality the MY property, and its counter-part being a joint characterization of the MGIG and Wishart distributions were proved in Massam and Wesołowski (2006). Here we will rephrase these results in a form adapted to the needs of the present paper.

First, let us define a probability distribution $W_{r,s}(q, c, a, b)$, by the density

$$f(x, y) \propto \begin{vmatrix} x & c \\ c^T & y \end{vmatrix}^{q - \frac{r+s+1}{2}} e^{-\langle a,x\rangle - \langle b,y\rangle} I_{\mathcal{M}}(x, y),$$

where $q \in \Lambda_{r+s}$, $q > \frac{r \wedge s - 1}{2}$, $a \in \mathcal{V}_r^+$, $b \in \mathcal{V}_s^+$, c is an $r \times s$ matrix of a full rank and

$$\mathcal{M} = \left\{(x, y) : \begin{pmatrix} x & c \\ c^T & y \end{pmatrix} \in \mathcal{V}_{r+s}^+\right\}.$$

The direct MY property for matrices of different dimension may be phrased as

Theorem 3 *If* $(X, Y) \sim W_{r,s}(q, c, a, b)$ *then*

$$(X, Y - c^T X^{-1} c) \sim \mathrm{MGIG}_r\left(q - \frac{s}{2}, a, cbc^T\right) \otimes W_s\left(q - \frac{r}{2}, b\right) \tag{7}$$

and

$$(X - cY^{-1}c^T, Y) \sim W_r\left(q - \frac{s}{2}, a\right) \otimes \mathrm{MGIG}_s\left(q - \frac{r}{2}, b, c^T ac\right). \tag{8}$$

The characterization has the form given in the following Theorem.

Theorem 4 *Let (X, Y) be a random vector taking values in $\mathcal{M}$ defined above for a given $r \times s$ matrix c of a full rank.*

If X and $Y - c^T X^{-1} c$ are independent and also Y and $X - cY^{-1}c^T$ are independent then there exist $q \in \Lambda_{r+s}$, $q > \frac{r \wedge s - 1}{2}$, $a \in \mathcal{V}_r^+$, $b \in \mathcal{V}_s^+$ such that $(X, Y) \sim W_{r,s}(q, c, a, b)$

In the sequel we will see how these result contribute towards further connections between the MGIG and Wishart matrices. This connection will be explored in Sects. 2, 3, and 4. In Sect. 5 the connection through a Wishart Bayesian model will be studied.

2 GIG mixtures and the Herz formula

We start with a recent observation by (V. Seshadri 2005, unpublished) who considered the distribution of a random vector (Y, T) as a GIG mixture, i.e. the conditional distribution of Y given T is $\mathrm{GIG}(\lambda, (2\alpha)^{-1}, \alpha T^2/2)$ with the mixing variable T having a Bessel type density

$$f_T(t) = \frac{t^{\nu-1}\mathcal{K}_\lambda(t)}{2^{\nu-2}\Gamma\left(\frac{\nu-\lambda}{2}\right)\Gamma\left(\frac{\nu+\lambda}{2}\right)} I_{(0,\infty)}(t),$$

where $\nu > \lambda$ is a parameter (here $\mathcal{K}_\lambda$ denotes the univariate modified Bessel function of the third kind). The joint density has the form

$$f_{(Y,T)}(y, t) = \frac{\alpha^{-\lambda} y^{\lambda-1} t^{\nu-\lambda-1}}{2^{\nu-1}\Gamma\left(\frac{\nu-\lambda}{2}\right)\Gamma\left(\frac{\nu+\lambda}{2}\right)} \exp\left(-\frac{y}{2\alpha} - \frac{\alpha t^2}{2y}\right) I_{(0,\infty)^2}(y, t). \tag{9}$$

It can be seen that Y and T^2/Y are independent gamma variables with parameters $((\nu+\lambda)/2, \alpha/2)$ and $((\nu-\lambda)/2, 1/(2\alpha))$, respectively, where the first element of the pair is the shape and the second is the scale parameter.

Note that in the case $\nu = \lambda + 1$ the above observation can be interpreted and much generalized while looking at properties of the Wishart distribution $W_n(p, a)$, $p > (n-1)/2$, when the matrix a is block-wise diagonal. To be more precise let $K = \begin{pmatrix} K_1, K_{12} \\ K_{21}, K_2 \end{pmatrix} \sim W_n(p, a)$ with the block decomposition of a as $a = \begin{pmatrix} a_1, 0 \\ 0, a_2 \end{pmatrix}$ (dimensions of the respective blocks of K and a are the same, say the diagonal blocks are $r \times r$ and $s \times s$, $r + s = n$). Then, it is well-known (see for instance Example 3.14 in Muirhead 1982) that K_1 and $K_{21}K_1^{-1}K_{12}$ are independent Wishart matrices (as a matter of fact jointly independent and Wishart are K_1, $K_{21}K_1^{-1}K_{12}$ and $K_{2\cdot1}$). Moreover the density of (K_1, K_{12}) is

$$f_{K_1,K_{12}}(k_1, k_{12}) = \frac{|a_1|^p |a_2|^{\frac{r}{2}} \Gamma_s\left(p - \frac{r}{2}\right)}{\Gamma_n(p)} |k_1|^{p-\frac{n+1}{2}} e^{-\langle a_1, k_1\rangle - \langle a_2, k_{12}^t k_1^{-1} k_{12}\rangle} \times I_{\mathcal{V}_r^+ \times \mathcal{M}_{r,s}}(k_1, k_{12}), \tag{10}$$

where $\mathcal{M}_{r,s}$ is the space of $r \times s$ real matrices. Integrating out k_1 we get the marginal density of K_{12} as

$$f_{K_{12}}(k_{12}) = \frac{|a_1|^p |a_2|^{\frac{r}{2}} \Gamma_s\left(p - \frac{r}{2}\right)}{\Gamma_{r+s}(p)} \int\limits_{\mathcal{V}_r^+} |k_1|^{p-\frac{s}{2}-\frac{r+1}{2}} e^{-\langle a_1, k_1\rangle - \langle k_{12}a_2k_{12}^T, k_1-1\rangle}\, dk_1$$
$$= \frac{|a_1|^p |a_2|^{\frac{r}{2}}}{\pi^{\frac{rs}{2}} \Gamma_r(p)} \mathcal{K}^{(r)}_{p-\frac{s}{2}}(a_1, k_{12}a_2k_{12}^T)$$

for all $k_{12} \in \mathcal{M}_{r,s}$, where the upper index (r) of the Bessel function means that the arguments are in $\mathcal{V}_r^+$.

Dually, we consider the joint distribution of (K_2, K_{12}), Similarly as above we obtain another formula for the density of K_{12}

$$f_{K_{12}}(k_{12}) = \frac{|a_1|^{\frac{s}{2}} |a_2|^p}{\pi^{\frac{rs}{2}} \Gamma_s(p)} \mathcal{K}^{(s)}_{p-\frac{r}{2}}(a_2, k_{12}^T a_1 k_{12})\, I_{\mathcal{M}_{r,s}}(k_{12}).$$

Comparing both expressions for the density of K_{12} we get the Herz identity for Bessel functions of arguments of different dimension.

Theorem 5 *Let $a \in \mathcal{V}_r^+$, $b \in \mathcal{V}_s^+$ and let $c \in \mathcal{M}_{r,s}$ be of full rank. Let $p > \frac{r-1}{2}$ and $q = p - \frac{r-s}{2}$. Then*

$$\Gamma_s(q)|a|^p \mathcal{K}_p^{(r)}(a, cbc^T) = \Gamma_r(p)|b|^q \mathcal{K}_q^{(s)}(b, c^T ac). \tag{11}$$

An extended version of (11), proved in Massam and Wesołowski (2006), was an essential tool in proving Theorem 3.

3 A characterization of a class of 2 × 2 Wishart matrices

The characterization related to MY property in the univariate case (see Letac and Wesołowski 2000) was used by Letac and Massam (2001) in their paper devoted to generalization of the result by Geiger and Heckerman (1998), and concerned with characterization of, so-called, quasi-Wishart distribution of a 2×2 symmetric positive definite random matrix $K = \begin{pmatrix} K_1, K_{12} \\ K_{21}, K_2 \end{pmatrix}$. We say that K has quasi-Wishart distribution if the conditional density of (K_1, K_2) given $K_{12} = k_{12}$ is

$$f_{(K_1,K_2)|K_{12}=k_{12}}(k_1, k_2) \propto (k_1k_2 - k_{12}^2)^{p-3/2} e^{-a_1k_1 - a_2k_2}$$

for $k_1, k_2 > 0$, $k_1k_2 > k_{12}^2$ and for any $k_{12} \in \mathbf{R}$. This distribution was characterized by the conditions of independences of (K_1, K_{12}) and $K_2 - K_{21}K_1^{-1}K_{12} = K_{2\cdot 1}$ and of (K_2, K_{21}) and $K_1 - K_{12}K_2^{-1}K_{21} = K_{1\cdot 2}$ in the two papers mentioned above. Also it was pointed out there that for 2×2 matrices the above independence conditions do not characterize the Wishart distribution.

In this section we are going to show that such a characterization is possible when the independence conditions are somewhat strengthened. This will lead us to special Wishart law with diagonal matrix variate parameter.

It is worth pointing out that higher-dimensional version of this result was proved in Geiger and Heckerman (2002). For $n \geq 3$ they proved that independences of (K_1, K_{12}) and $K_{2\cdot 1}$ and of (K_2, K_{21}) and $K_{1\cdot 2}$ for all possible block partitions of the symmetric positive definite random matrix $K = \begin{pmatrix} K_1, K_{12} \\ K_{21}, K_2 \end{pmatrix}$ imply that K is Wishart. This result has been improved recently in Massam and Wesołowski (2006), where only three pairs of the independence conditions were assumed. The main tool was the characterization related to MY property for matrices of different dimensions, see Theorem 4.

Recall, that for a 2×2 Wishart $W_2(p, c)$ matrix $K = \begin{pmatrix} K_1 & K_{12} \\ K_{12} & K_2 \end{pmatrix}$ with diagonal c the variables

$$K_1,\ K_{12}^2/K_1,\ K_{2\cdot 1} \text{ are independent} \tag{12}$$

and

$$K_2,\ K_{12}^2/K_2,\ K_{1\cdot 2} \text{ are independent.} \tag{13}$$

We will show that the above two conditions characterize a family of distributions on $\mathcal{V}_2^+$ including Wishart. To get a characterization of the Wishart family alone we need to specify some additional properties besides (12) and (13).

Theorem 6 *Let K be a random matrix valued in $\mathcal{V}_2^+$. If the independence conditions (12) and (13) hold then there exist positive numbers p, q, a, b such that the random vector (K_1, K_2, K_{12}^2) has the density f of the form*

$$f(k_1, k_2, k_{12}^2) = \frac{(ab)^{p+q}}{\Gamma(p+q)\Gamma(p)\Gamma(q)} (k_{12}^2)^{q-1}(k_1k_2 - k_{12}^2)^{p-1} e^{-ak_1 - bk_2} \\ I_{\mathcal{C}}(k_1, k_2, k_{12}^2), \tag{14}$$

where $\mathcal{C} = \{(k_1, k_2, k_{12}^2) : k_1k_2 > k_{12}^2,\ k_1, k_2 > 0\}$.

Proof Denote

$$X = K_1, \quad Y = \frac{K_{12}^2}{K_1} \quad \text{and} \quad Z = K_{2\cdot 1}.$$

Then

$$K_2 = Y + Z, \quad \frac{K_{12}^2}{K_2} = \frac{XY}{Y+Z} \quad \text{and} \quad K_{1\cdot 2} = \frac{ZX}{Y+X}.$$

Consequently (13) implies that

$$Y + Z \quad \text{and} \quad \frac{K_{12}^2}{K_2}/K_{1\cdot 2} = \frac{Y}{Z}$$

are independent. Note that this is the very moment the dimensionality 2×2 of the problem is crucial. Then, the classical Lukacs characterization of the gamma law (see Lukacs 1955) implies that Y and Z have gamma distributions with the same scale parameter, say $G(p, b)$ and $G(q, b)$. Thus $Y + Z = K_2$ is also gamma $G(p + q, b)$.

By symmetry, it follows that $\frac{K_{12}^2}{K_2}$ and $K_{1\cdot 2}$ have gamma distributions $G(\tilde{p}, a)$ and $G(\tilde{q}, a)$ and similarly as for K_2 we conclude that $X = K_1$ is gamma $G(\tilde{p} + \tilde{q}, a)$.

Consequently, the joint density of (K_1, K_2, K_{12}^2) can be derived from the joint density of (X, Y, Z) which is a random vector with independent gamma components. Since the transformation $(k_1, k_2, k_{12}^2) \to (k_1, k_{12}^2/k_1, k_{2\cdot 1})$ is a bijection from $\mathcal{C}$ onto $(0, \infty)^3$ with the jacobian of the inverse equal to $1/k_1$ we get on $\mathcal{C}$

$$\begin{aligned} f(k_1, k_2, k_{12}^2) &= \frac{1}{k_1} \frac{a^{\tilde{p}+\tilde{q}}}{\Gamma(\tilde{p}+\tilde{q})} k_1^{\tilde{p}+\tilde{q}-1} e^{-ak_1} \frac{b^p}{\Gamma(p)} \left(\frac{k_{12}^2}{k_1}\right)^{p-1} \\ &\quad \times e^{-b\frac{k_{12}^2}{k_1}} \frac{b^q}{\Gamma(q)} \left(k_2 - \frac{k_{12}^2}{k_1}\right)^{q-1} e^{-b\left(k_2 - \frac{k_{12}^2}{k_1}\right)} \\ &= \frac{a^{\tilde{p}+\tilde{q}} b^{p+q}}{\Gamma(\tilde{p}+\tilde{q})\Gamma(p)\Gamma(q)} k_1^{\tilde{p}+\tilde{q}-p-q} \left(k_{12}^2\right)^{p-1} \\ &\quad \times (k_1 k_2 - k_{12}^2)^{q-1} e^{-ak_1 - bk_2}. \end{aligned} \tag{15}$$

Dually, on $\mathcal{C}$ we have

$$f(k_1, k_2, k_{12}^2) = \frac{a^{\tilde{p}+\tilde{q}} b^{p+q}}{\Gamma(p+q)\Gamma(\tilde{p})\Gamma(\tilde{q})} k_2^{p+q-\tilde{p}-\tilde{q}} \left(k_{12}^2\right)^{\tilde{p}-1} (k_1 k_2 - k_{12}^2)^{\tilde{q}-1} e^{-ak_1 - bk_2}.$$

Thus $\tilde{p} = p$ and $\tilde{q} = q$ and finally we get (14). □

Remark 1 Note that the converse of Theorem 6 follows rather immediately by reverse reading of (15) with $\tilde{p} = p$ and $\tilde{q} = q$. That is: if a random vector (K_1, K_2, K_{12}^2) has the density (14) then

$$\left(K_1,\ \frac{K_{12}^2}{K_1},\ K_{2\cdot 1}\right) \sim G(p+q, a) \otimes G(p, b) \otimes G(q, b)$$

and

$$\left(K_2,\ \frac{K_{12}^2}{K_2},\ K_{1\cdot 2}\right) \sim G(p+q, b) \otimes G(p, a) \otimes G(q, a)$$

To identify the Wishart distribution we impose some moment and symmetry conditions additionally.

Corollary 1 *Let K be a random matrix in $\mathcal{V}_2^+$ such that (12) and (13) are satisfied. Assume that*

$$\mathrm{Var}\left(\frac{K_{12}^2}{K_i}\right) = 2E\left(\frac{K_{12}^2}{K_i}\right) \tag{16}$$

either for $i = 1$ or for $i = 2$ and that $(K_1, K_2, K_{12}) \stackrel{d}{=} (K_1, K_2, -K_{12})$. Then K is Wishart $W_2(Q, c)$, where c is a diagonal matrix and $Q > 1/2$.

Proof Note that the symmetry condition implies that the density f_K of the matrix K can be written in terms of the density f of the random vector (K_1, K_2, K_{12}^2) as

$$f_K(k_1, k_2, k_{12}) = |k_{12}| f(k_1, k_2, k_{12}^2).$$

Moreover (16) implies that $p = 1/2$ in (14), which together with the above gives

$$f_K(k_1, k_2, k_{12}) = \frac{(ab)^{q+\frac{1}{2}}}{\Gamma(q+1/2)\Gamma(q)\sqrt{\pi}}(k_1 k_2 - k_{12}^2)^{q-1} e^{-ak_1 - bk_2}$$

for $k_1 k_2 > k_{12}^2$, $k_1 > 0$. Note that (Theorem 2.1.12 in Muirhead 1982) $\Gamma_2(Q) = \Gamma(Q)\Gamma(Q-1/2)\sqrt{\pi}$. Consequently, K is Wishart $W_2(Q, c)$ with $Q = q + 1/2$ and diagonal c, with $c_{11} = a$ and $c_{22} = b$. □

4 Linear transformations of MGIGs

It is well known that if X is Wishart, $W_n(p, a)$, and M is a $k \times n$ constant matrix of rank k then

$$MXM^T \sim W_k\left(p, (Ma^{-1}M^T)^{-1}\right) \tag{17}$$

(see Theorem 3.2.5 in Muirhead 1982) and for $p > \frac{n-1}{2}$

$$(MX^{-1}M^T)^{-1} \sim W_k\left(p - \frac{n-k}{2}, MaM^T\right), \tag{18}$$

see Theorem 3.2.11 in Muirhead (1982).

In this section we present analogues of these results for the MGIG distribution. In this context the MY property of Theorem 3 will be very useful.

Theorem 7 *Let Y be a random matrix having the* $\mathrm{MGIG}_n\left(p, M^T aM, b\right)$ *distribution, where M is a* $k \times n$ *matrix of rank k,* $a \in \mathcal{V}_k^+$ *and* $b \in \mathcal{V}_n^+$*. Then*

$$MYM^T \sim \mathrm{MGIG}_k\left(p + \frac{n-k}{2}, a, MbM^T\right). \tag{19}$$

Let Z be a random matrix having the $\mathrm{MGIG}_n\left(q, b, M^T aM, \right)$ *distribution, where M is a* $k \times n$ *matrix of rank k,* $a \in \mathcal{V}_k^+$ *and* $b \in \mathcal{V}_n^+$*. Then*

$$(MZ^{-1}M^T)^{-1} \sim \mathrm{MGIG}_k\left(q - \frac{n-k}{2}, MbM^T, a\right).$$

Proof Note that the first part follows from the second by taking $Y = Z^{-1}$ via (2). Consider Z as a marginal of $(Z, V) \sim W_{n,k}(Q, M, b, a)$ with $Q = q + \frac{k}{2}$. By Theorem 3 it follows that

$$E\left(e^{\langle \theta, V\rangle}\right) = E\left(e^{\langle \theta, V - MZ^{-1}M^T\rangle}\right) E\left(e^{\langle \theta, MZ^{-1}M^T\rangle}\right).$$

Moreover, since the distributions of V and $V - MZ^{-1}M^T$ are known from the above equality, by (4), we get

$$E\left(e^{\langle \theta, MZ^{-1}M^T\rangle}\right) = \frac{\mathcal{K}^{(k)}_{Q-\frac{n}{2}}(a-\theta, MbM^T)}{\mathcal{K}^{(k)}_{Q-\frac{n}{2}}(a, MbM^T)} \frac{|a|^{Q-\frac{n}{2}}}{|a-\theta|^{Q-\frac{n}{2}}}.$$

Since $\mathcal{K}_p(\alpha, \beta)|\beta|^p = \mathcal{K}_p(\beta, \alpha)|\alpha|^p$, which is a special version of (11), we get

$$E\left(e^{\langle \theta, MZ^{-1}M^T\rangle}\right) = \frac{\mathcal{K}^{(k)}_{Q-\frac{n}{2}}(MbM^T, a-\theta)}{\mathcal{K}^{(k)}_{Q-\frac{n}{2}}(MbM^T, a)}.$$

Again using (4) we get the conclusion. □

Note that the second result of Theorem 7 is a straightforward generalization of (18) by taking $a \to 0$ in the definition of the distribution of Z. On the other hand, though the conclusion of the first part of Theorem 7 looks similar to (17), essentially these two are different.

5 MGIG as a conjugate prior for Wishart family

The original definition of the MGIG distribution, due to Barndorff-Nielsen et al. (1982) was in the context of exponential transformation models. Below we show that MGIG can be defined also as the posterior distribution in a Wishart Bayesian model with a Wishart prior.

Proposition 1 *Let (X, Y) be a pair of random matrices valued in $\mathcal{V}_n^+ \times \mathcal{V}_n^+$ such that the conditional distribution of Y given X is Wishart, i.e. $Y|X \sim W_n(p, X^{-1})$, $p > \frac{n-1}{2}$ and the marginal distribution of X is Wishart, $W_n(p + r, a)$, where $p + r > \frac{n-1}{2}$, $a \in \mathcal{V}_n^+$. Then the conditional distribution of X given Y is MGIG,*

$$X|Y \sim \mathrm{MGIG}_n(r, a, Y).$$

We specified the above result only due to the fact that it gives a natural definition of MGIG through Wishart, which is different than that which is hidden in Theorem 3. Nevertheless a more general result can be proved.

Theorem 8 *Let (X, Y) be a pair of random matrices valued in $\mathcal{V}_n^+ \times \mathcal{V}_n^+$ such that the conditional distribution of Y given X is Wishart, i.e. $Y|X \sim W_n(p, X^{-1})$, $p > \frac{n-1}{2}$ and the marginal distribution of X is $\mathrm{MGIG}_n(p + r, a, b)$, where $a, b, \in \mathcal{V}_n^+$. Then the conditional distribution of X given Y is MGIG,*

$$X|Y \sim \mathrm{MGIG}_n(r, a, b + Y).$$

Proof The computation is standard. First we find the marginal density of Y:

$$\begin{aligned} f_Y(y) &= \int_{\mathcal{V}_n^+} \frac{|y|^{p-\frac{n+1}{2}}}{\Gamma_n(p)|x|^p} e^{-\langle x^{-1}, y\rangle} \frac{|x|^{p+r-\frac{n-1}{2}}}{\mathcal{K}_{p+r}(a,b)} e^{-\langle a,x\rangle - \langle b, x^{-1}\rangle}\, dx \\ &= \frac{|y|^{p-\frac{n+1}{2}}}{\Gamma_n(p)\mathcal{K}_{p+r}(a,b)} \int_{\mathcal{V}_n^+} |x|^{r-\frac{n-1}{2}} e^{-\langle a,x\rangle - \langle b+y, x^{-1}\rangle}\, dx \\ &= \frac{|y|^{p-\frac{n+1}{2}} \mathcal{K}_r(a, b+y)}{\Gamma_n(p)\mathcal{K}_{p+r}(a,b)} \end{aligned} \tag{20}$$

for $y \in \mathcal{V}_n^+$. Now, the conditional density for $x, y \in \mathcal{V}_n^+$ is

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f(x,y)}{f_Y(y)} = \frac{\frac{|y|^{p-\frac{n+1}{2}}}{\Gamma_n(p)\mathcal{K}_{p+r}(a,b)} |x|^{r-\frac{n-1}{2}} e^{-\langle a,x\rangle - \langle b+y, x^{-1}\rangle}}{\frac{|y|^{p-\frac{n+1}{2}}\mathcal{K}_r(a,b+y)}{\Gamma_n(p)\mathcal{K}_{p+r}(a,b)}} \\ &= \frac{|x|^{r-\frac{n+1}{2}} e^{-\langle a,x\rangle - \langle b+y, x^{-1}\rangle}}{\mathcal{K}_r(a, b+y)} \end{aligned} \tag{21}$$

□

The above theorem says that MGIG is a conjugate prior for the Wishart family—see Diaconis and Ylvisaker (1979).

Remark 2 The Bayesian Wishart–MGIG model defined above has an interesting independence property which is inherited from a wider model. Namely if (X, Y) are such that Y given X is Wishart, i.e. $Y|X \sim W_n(p, X^{-1})$, $p > \frac{n-1}{2}$, then $U = X^{-1/2} Y X^{-1/2}$ and X are independent and U is Wishart, $W_n(p, e)$, where $e \in \mathcal{V}_n^+$ is the identity matrix. It follows immediately from (17) with $M = X^{-1/2}$ and $a = X^{-1}$.

Remark 3 Note that looking at the Bessel-like density of the random matrix Y in the model we consider in this section—see the proof of Theorem 8—we get an identity for integrals of the matrix variate Bessel function

$$\int_{\mathcal{V}_n^+} |y|^{p-\frac{n+1}{2}} \mathcal{K}_r(a, b+y)\, dy = \Gamma_n(p) \mathcal{K}_{p+r}(a, b) \tag{22}$$

for any $p > \frac{n-1}{2}$ and $a, b \in \mathcal{V}_n^+$.

Maybe it is worth specifying to special cases: first we take $b \to 0$ in (22) getting

$$\int_{\mathcal{V}_n^+} |y|^{p-\frac{n+1}{2}} \mathcal{K}_r(a, y)\, dy = \frac{\Gamma_n(p)\Gamma_n(p+r)}{|a|^{p+r}}$$

and the simplest one with $p = \frac{n+1}{2}$ gives

$$\int_{\mathcal{V}_n^+} \mathcal{K}_r(a, y)\, dy = \frac{\Gamma_n\left(\frac{n+1}{2}\right)\Gamma_n\left(\frac{n+1}{2}+r\right)}{|a|^{\frac{n+1}{2}+r}}.$$

Remark 4 Dually to Theorem 8 we have the following result: If $Y|X \sim W_n(p, X)$, $p > \frac{n-1}{2}$ and $X \sim \mathrm{MGIG}_n(r-p, a, b)$, where $a, b, \in \mathcal{V}_n^+$. Then

$$X|Y \sim \mathrm{MGIG}_n(r, a+Y, b).$$

Next we will see that the Wishart–MGIG model can be uniquely specified by one conditional distribution and one regression condition. Such conditional specifications of statistical models were considered by many authors, for a comprehensive review of results and references, see for instance Chap. 7 in Arnold et al. (1999).

Theorem 9 *Assume that* $X|Y \sim \mathrm{MGIG}_n(r, a, b+Y)$ *and* $E(Y|X) = pX$*, where* $a, b \in \mathcal{V}_n^+$, $p, p+r > \frac{n-1}{2}$.

Then $Y|X \sim W_n(p, X^{-1})$ *and* $X \sim \mathrm{MGIG}_n(p+r, a, b)$.

Proof By the generalized Bayes rule we have

$$px \int_{\mathcal{V}_n^+} f_{X|Y=y}(x) f_Y(y)\, dy = \int_{\mathcal{V}_n^+} y f_{X|Y=y}(x) f_Y(y)\, dy, \quad \text{a.e. in } \mathcal{V}_n^+.$$

Since $X|Y \sim \mathrm{MGIG}_n(r, a, b+Y)$, changing x to x^{-1}, we get

$$px^{-1} \int_{\mathcal{V}_n^+} e^{yx} \frac{f_Y(y)}{\mathcal{K}_r(a, b+y)}\, dy = \int_{\mathcal{V}_n^+} ye^{yx} \frac{f_Y(y)}{\mathcal{K}_r(a, b+y)}\, dy.$$

Denote by L the Laplace transform of $\frac{f_Y(y)}{K_r(a,b+y)}$. Then the above equation reads

$$px^{-1}L(x) = L'(x).$$

Due to the fact that the derivative of $\log(|x|)$ with respect to x is x^{-1}, from the above eqnarray we get $L(x) = N|x|^p$, where N is a constant. By the uniqueness of the Laplace transform we get

$$\frac{f_Y(y)}{\mathcal{K}_r(a, b+y)} = C|y|^{p-\frac{n+1}{2}},$$

where C is a constant. Now, the form of the constant C follows from (22), and thus f_Y is defined as in (20). Since the conditional distribution $X|Y$ is given the model described in Theorem 8 is uniquely identified. □

The Wishart–MGIG model considered in Theorem 8 can be further expanded into MGIG–MGIG model as follows.

Remark 5 Let

$$Y|X \sim \mathrm{MGIG}_n\left(p,\ \lambda c^{1/2}Xc^{1/2},\ \mu(c^{1/2}Xc^{1/2})^{-1}\right), \tag{23}$$

where λ, μ are non-negative numbers and $c \in \mathcal{V}_n^+$ and $X \sim \mathrm{MGIG}_n(q, a, b)$. Then the density of Y is

$$f_Y(y) = \frac{|c|^p|y|^{p-\frac{n+1}{2}}}{\mathcal{K}_p(\lambda e, \mu e)\mathcal{K}_q(a, b)} \mathcal{K}_{p+q}(a + \lambda c^{1/2}yc^{1/2}, b + \mu(c^{1/2}yc^{1/2})^{-1})\ I_{\mathcal{V}_n^+}(y)$$

and

$$X|Y \sim \mathrm{MGIG}_n(p+q, a + \lambda c^{1/2}yc^{1/2}, b + \mu(c^{1/2}yc^{1/2})^{-1}).$$

Similarly as in the Wishart–MGIG model, the condition (23) alone implies, through (19), that X and $U = X^{1/2}c^{1/2}Yc^{1/2}X^{1/2}$ are independent and $U \sim \mathrm{MGIG}_n(p, \lambda e, \mu e)$.
Furthermore the integral identity takes the form: for $p > \frac{n-1}{2}$

$$\int_{\mathcal{V}_n^+} |y|^{p-\frac{n+1}{2}}\ \mathcal{K}_{p+q}\left(a + \lambda c^{1/2}yc^{1/2}, b + \mu(c^{1/2}yc^{1/2})^{-1}\right)\, dy = \frac{\mathcal{K}_p(\lambda e, \mu e)\mathcal{K}_q(a, b)}{|c|^p}.$$

Let us point out to the fact that Seshadri (2003), extending the univariate result of Vallois (1989), obtained the following result: If $X \sim W_n(p, a)$ and $Y \sim W_n(q, e)$ are independent, $p, q > \frac{n-1}{2}$ and $U = X^{1/2}YX^{1/2}$ then $X|U \sim \mathrm{MGIG}_n\left(\frac{p-q}{2}, a, U\right)$, which can be regarded as another definition of the MGIG distribution.

Acknowledgments The authors are indebted to the referee for many helpful suggestions. They led to many improvements, in particular, to improved versions of Remarks 2 and 5.

References

Arnold BC, Castillo E, Sarabia J-M (1999) Conditional specification of statistical models. Springer, New York

Barndorff-Nielsen O, Blæsild P, Jensen JL, Jørgensen B (1982) Exponential transformation models. Proc R Soc Lond A 379:41–65

Bernadac E (1995) Random continued fractions and inverse gaussian distribution on a symmetric cone. J Theor Probab 8:221–259

Butler RW (1998) Generalized inverse Gaussian distributions and their Wishart connections. Scand J Stat 25:69–75

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

Diaconis P, Ylvisaker D (1979) Conjugate priors for exponential families. Ann Stat 7:269–281

Geiger D, Heckerman D (1998) A characterization of the bivariate Wishart distribution. Probab Math Stat 18:119–131

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

Gindikin SG (1975) Invariant generalized functions in homogeneous domains. Funct Anal Appl 9:50–52

Herz CS (1955) Bessel functions of matrix argument. Ann Math 61:474–523

Letac G (2003) Convergence de l'intégrale GIG, unpublished note, pp 1–2

Letac G, Massam H (2001) The normal quasi-Wishart distribution. In: Viana MAG, Richards DStP (eds) Algebraic methods in statistics and probability. AMS Contemp Math 287:231–239

Letac G, Seshadri V (1983) A characterization of the generalized inverse Gaussian distribution by continued fractions. Z Wahrsch Verw Geb 62:482–489

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

Lukacs E (1955) A characterization of the gamma distribution. Ann Math Stat 26:319–324

Massam H, Wesołowski J (2006) The Matsumoto-Yor property and the structure of the Wishart distribution. J Multivar Anal 97:103–123

Muirhead RJ (1982) Aspects of multivariate statistical theory. Wiley, New York

Seshadri V (2003) Some properties of the matrix generalized inverse Gaussian distribution. In: Balakrishnan N, Kannan N, Srinivasan MR (eds) Statistical methods and practice. Recent advances. Narosa Publishing House, New Delhi pp 47–56

Vallois P (1989) Sur le passage de certaines marches aléatoires planes au-dessus d'une hyperbole équilatère. Ann Inst H Poincaré 25:443–456

Wesołowski J (2002) The Matsumoto-Yor independence property for GIG and Gamma laws, revisited. Math Proc Camb Philos Soc 133:153–161