Electronic Journal of Statistics
Vol. 5 (2011) 146–158
ISSN: 1935-7524
DOI: 10.1214/11-EJS602

On the mean and variance of the generalized inverse of a singular Wishart matrix*

R. Dennis Cook†

School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455
e-mail: dennis@stat.umn.edu

and

Liliana Forzani

Instituto de Matemática Aplicada del Litoral and Facultad de Ingeniería Química, CONICET and UNL, Güemes 3450, (3000) Santa Fe, Argentina
e-mail: liliana.forzani@gmail.com

Abstract: We derive the first and the second moments of the Moore-Penrose generalized inverse of a singular standard Wishart matrix without relying on a density. Instead, we use the moments of an inverse Wishart distribution and an invariance argument which is related to the literature on tensor functions. We also find the order of the spectral norm of the generalized inverse of a Wishart matrix as its dimension and degrees of freedom diverge.

AMS 2000 subject classifications: Primary 62H05, secondary 62E15.
Keywords and phrases: Inverse Wishart distribution, Moore-Penrose generalized inverse, singular inverse Wishart distributions, tensor functions.

Received October 2010.

Contents

*The authors thank Joe Eaton and a referee for helpful suggestions.
†This work was supported in part by grant DMS-1007547 from the U.S. National Science Foundation.

1. Introduction

Let the columns of $X = (X_1, \ldots, X_n) \in \mathcal{R}^{p \times n}$ be a sample X_i, $i = 1, \ldots, n$, from $N_p(0, \Sigma)$, the p-variate normal distribution with mean 0 and positive definite variance matrix Σ. The sum of squares matrix $S = XX^T$ follows the p-variate Wishart distribution with n degrees of freedom, denoted by $W_p(\Sigma, n)$. If $p > n$, then $W_p(\Sigma, n)$ is called the singular Wishart distribution. Its density was given by Uhlig [16] under the Hausdorff measure and by Srivastava [14] under the Lebesgue measure on the functionally independent elements of S.

Let $W \sim W_p(\Sigma, n)$ and let W^- be the usual Moore-Penrose inverse of W, defined as the unique matrix W^- such that $W^-WW^- = W^-$, $WW^-W = W$, and WW^- and W^-W are symmetric. As indicated by this definition, we use the usual inner products based on the identity matrix I_p for the two symmetry conditions in the Moore-Penorse inverse, unless stated otherwise. If W is non-singular then W^- is the regular inverse. The distribution of W^- is called the inverse Wishart distribution when $p \leq n$ and the generalized inverse Wishart distribution when $p > n$. Díaz-García and Gutiérrez-Jáimez [4] gave an expression for the density function of the generalized inverse Wishart distribution under the Hausdorff measure. Under the Lebesgue measure on the functionally independent elements, the density was proposed by Bodnar and Okhrin [2] and Zhang [18] but their results seem inconsistent. The density given by Bodnar and Okhrin involves the eigenvalues of W^-, while the density given by Zhang does not. Both of their results were based on the density of the singular Wishart distribution given by Srivastava [14], and neither gave moments of the distribution. If $W \sim W_p(\Sigma, n)$ and $p \leq n$, then we denote the inverse Wishart distribution of W^{-1} as $W_p^{-1}(\Sigma, n)$. If $p > n$ the distribution of W^- is denoted as $W_p^-(\Sigma, n)$.

In this note we derive the first two moments of $W_p^-(I_p, n)$ without relying on an expression of its density function, and discuss the issues involved in extending this result to $W_p^-(\Sigma, n)$. Our results are based on the first two moments of the inverse Wishart distribution [17] plus an invariance argument. We also find the order of the spectral norm $\|\cdot\|$ of $W_p(\Sigma, n)$ as $n, p \to \infty$. In addition to being a contribution to random matrix theory, these results may play a role in Bayesian analysis because the corresponding distributions are natural conjugate priors for the covariance matrix in the normal distribution [4]. They are also useful in studies of estimation methods for high dimensional $n < p$ regressions.

We present our findings on the moments $W_p^-(I, n)$ in Theorem 2.1 of Section 2. Those findings rely on an invariance relationship that is described in Proposition 4.1 and is related to the classical mechanics literature on tensors. Results on the moments of a $W_p^-(\Sigma, n)$ random matrix and on its order are given in Section 3.2. The proof of Theorem 2.1 is given in Section 4 and the proof of Proposition 4.1 is given in Section 4.5.

Throughout this article $\sim$ means equal in distribution and $\mathcal{R}^{p\times q}$ denotes the collection of all real $p\times q$ matrices. For sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ if there are constants m, M and N such that $0 < m < |a_n/b_n| < M < \infty$ for all $n > N$. The Kronecker product $\otimes$ of two matrices $A = (A_{ij}) \in \mathcal{R}^{a\times b}$ and $B \in \mathcal{R}^{c\times d}$ is the $ac \times bd$ matrix expressed in block form as $A \otimes B = (A_{ij}B)$, $i = 1,\ldots,a,\ j = 1,\ldots,b$. The vec operator [8] maps $A \in \mathcal{R}^{a\times b}$ to $\mathcal{R}^{ab}$ by stacking its columns. We use $e_i \in \mathcal{R}^p$ to denote the vector with a 1 in the i-th position and 0's elsewhere. The $p^2 \times p^2$ commutation matrix is denoted as $C_{p^2} = \sum_{i,j}(e_i \otimes e_j)(e_j^T \otimes e_i^T)$ [10, 11].

2. Moments of $W_p^-(I_p, n)$

The first two moments of $W_p^{-1}(I_p, n)$ have been known for some time:

Proposition 2.1. *(von Rosen, 1988) Let $W \sim W_p(I_p, n)$. If $n > p+3$, then rank$(W) = p$ with probability 1,*

$$\begin{aligned} E(W^{-1}) &= a_1 I_p, \\ E(W^{-1}W^{-1}) &= b_1 I_p, \\ var\{vec(W^{-1})\} &= c_1(I_{p^2} + C_{p^2}) + 2d_1\, vec(I_p) vec^T(I_p), \end{aligned}$$

where $a_1 = (n-p-1)^{-1}$, $b_1 = (n-1)c_1$, $c_1^{-1} = (n-p)(n-p-1)(n-p-3)$, and $d_1^{-1} = (n-p)(n-p-1)^2(n-p-3)$.

There is a close relationship between the form of var$\{$vec$(W^{-1})\}$ and the spectral decomposition of a fourth-order isotropic tensor from classical elasticity theory: Expressing var$\{$vec$(W^{-1})\}$ in terms of the elements of W_{ij}^{-1} of W^{-1} and rearranging terms we have

$$\begin{aligned} \mathrm{cov}(W_{ij}^{-1}, W_{kl}^{-1}) &= c_1(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + 2d_1\delta_{ij}\delta_{kl} \\ &= \{(2c_1 + 6d_1)/3\}\delta_{ij}\delta_{kl} + 2c_1\{(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2 - \delta_{ij}\delta_{kl}/3\}, \end{aligned}$$

where Kronecker's delta $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Except for the coefficients $(2c_1 + 6d_1)$ and $2c_1$, this form is identical to a classical tensor decomposition that is related to the bulk and shear moduli (see, for example, equation (14) of [1] and the associated references). Further comments on the relationship between this work and continuum mechanics are given in Section 4. von Rosen (1988) gives also various moments of $W_p^{-1}(\Sigma, n)$, but here our focus is on Wisharts with $\Sigma = I_p$.

The following theorem gives results for $W_p^-(I_p, n)$ that are analogous to those stated in Proposition 2.1 for $W_p^{-1}(I_p, n)$.

Theorem 2.1. *Let $W \sim W_p(I_p, n)$. If $p > n+3$, then rank$(W) < p$ with probability 1,*

$$\begin{aligned} E(W^-) &= a_2 I_p, \\ E(W^-W^-) &= b_2 I_p, \end{aligned}$$

$$\begin{aligned} var\{vec(W^-)\} &= c_2(I_{p^2} + C_{p^2}) + 2d_2 vec(I_p)vec^T(I_p), \\ var\{tr(W^-)\} &= 2c_3 n + 2d_3 n^2, \end{aligned}$$

where

$$\begin{aligned} a_2 &= \frac{n}{p(p-n-1)}, \\ b_2 &= \frac{n(p-1)c_3}{p}, \\ c_2 &= \frac{n\{p(p-1) - n(p-n-2) - 2\}c_3}{p(p-1)(p+2)}, \\ d_2 &= \frac{n\{n^2(n-1) + 2n(p-2)(p-n) + 2p(p-1)\}d_3}{p^2(p-1)(p+2)}, \end{aligned}$$

$c_3^{-1} = (p-n)(p-n-1)(p-n-3)$ *and* $d_3^{-1} = (p-n)(p-n-1)^2(p-n-3)$.

The constraints $n > p+3$ and $p > n+3$ in Proposition 2.1 and Theorem 2.1 are needed to ensure that the moments exist. Comparing Proposition 2.1 and Theorem 2.1 we see that the first two moments of W^- have the same functional form in the singular and nonsingular cases, differing only by a_j, b_j, c_j and d_j, $j = 1, 2$. The following corollary gives the asymptotic magnitudes of these factors. Essentially, it tells us that their asymptotic behavior depends weakly on the rank of W. Its proof seems straightforward and is omitted.

Corollary 2.1. *If* $n > p+3$ *and* $p/n \to r$ *with* $0 \leq r < 1$ *as* $p, n \to \infty$ *then* $a_1 \asymp n^{-1}$, $b_1 \asymp n^{-2}$, $c_1 \asymp n^{-3}$ *and* $d_1 \asymp n^{-4}$. *If* $p > n+3$ *and* $n/p \to r$ *with* $0 < r < 1$ *as* $p, n \to \infty$ *then* $a_2 \asymp p^{-1} \asymp n^{-1}$, $b_2 \asymp p^{-2} \asymp n^{-2}$, $c_2 \asymp p^{-3} \asymp n^{-3}$ *and* $d_2 \asymp p^{-4} \asymp n^{-4}$.

To gain some intuition about the structure of the variance in Theorem 2.1, we first recognize that one role of $P_s \equiv (I_{p^2} + C_{p^2})/2$ is to project onto the space of symmetric $p \times p$ matrices: Let A be a $p \times p$ matrix. Then $P_s \text{vec}(A) = \{\text{vec}(A) + \text{vec}(A^T)\}/2$ and $\text{vec}^{-1}(P_s A) = (A + A^T)/2$. Also, $P_v \equiv \text{vec}(I_p)\text{vec}^T(I_p)/p$ projects onto $\text{span}\{\text{vec}(I_p)\}$. Consequently,

$$\begin{aligned} \text{var}\{\text{vec}(W^-)\} &= 2c_2 P_s + 2pd_2 P_v \\ &= 2c_2(P_s - P_v) + 2(c_2 + pd_2)P_v, \end{aligned}$$

where $P_s - P_v$ and P_v are orthogonal projection operators. One implication of this development is given in the following corollary.

Corollary 2.2. *The eigenvalues of* $vec\{var(W^-)\}$ *are* $2(c_2 + pd_2)$ *with multiplicity 1,* $2c_2$ *with multiplicity* $p-1$ *and 0 with multiplicity* $p(p-1)$. *The corresponding eigenvectors are* $vec(I_p)/\sqrt{p}$, *the eigenvectors of* $P_s - P_v$ *and the eigenvectors of* $I_{p^2} - P_s$.

If $W \sim W_p(\Sigma, n)$ then $\Sigma^{-1/2} W \Sigma^{-1/2} \sim W_p(I_p, n)$ and

$$W^\dagger = \Sigma^{-1/2}\big(\Sigma^{-1/2} W \Sigma^{-1/2}\big)^- \Sigma^{-1/2}$$

is a reflexive generalized inverse of W. It follows straightforwardly that

Corollary 2.3. *If $W \sim W_p(\Sigma, n)$ and $p > n + 3$ then*

$$\begin{aligned} E(W^\dagger) &= a_2 \Sigma^{-1} \\ var\{vec(W^\dagger)\} &= c_2(I_{p^2} + C_{p^2})(\Sigma^{-1} \otimes \Sigma^{-1}) + 2d_2\, vec(\Sigma^{-1}) vec^T(\Sigma^{-1}) \\ cov(W^\dagger_{ij}, W^\dagger_{kl}) &= c_2(\Sigma^{-1}_{ik}\Sigma^{-1}_{jl} + \Sigma^{-1}_{il}\Sigma^{-1}_{jk}) + 2d_2 \Sigma^{-1}_{ij}\Sigma^{-1}_{kl}, \end{aligned}$$

where $W^\dagger_{ij}$ and Σ^{-1}_{ij} denote element (i, j) of $W^\dagger$ and Σ^{-1}. The second and third conclusions are the same, except the third is in terms of the elements of $W^\dagger$.

The form of $\mathrm{var}\{\mathrm{vec}(W^\dagger)\}$ given in Corollary 2.3 is identical to the asymptotic variance of the covariance matrix from a sample from an elliptically contoured distribution. In that case the constants $c_2 = 1 + \kappa$ and $d_2 = \kappa/2$, where κ is the kurtosis of the distribution (see, for example, Tyler [15]).

Because $\Sigma W^\dagger W$ and $\Sigma^{-1} W W^\dagger$ are symmetric, the reflexive generalized inverse $W^\dagger$ of W is also the Moore-Penrose inverse in the inner products based on Σ and Σ^{-1} [13], but $W^\dagger$ it is not the usual Moore-Penrose inverse since $W^\dagger W$ and $W W^\dagger$ are not symmetric. We were unable to find succinct expressions for the mean and variance of $W^-(\Sigma, n)$ that are analogous to those for $W^-(I_p, n)$ given in Theorem 2.1. In the next section we give some results on the moments of $W_p^-(\Sigma, n)$. We also give the order of the spectral norm $\|\cdot\|$ of the scaled Wishart $\Sigma^{1/2} W_p^-(\Sigma, n) \Sigma^{1/2}$ as $n, p \to \infty$, which may be helpful in asymptotic studies of regressions with $p > n$.

3. Properties of $W_p^-(\Sigma, n)$

3.1. Mean and variance

Let $W \sim W_p(\Sigma, n)$ with $p > n + 3$. The singular Wishart matrix W can be decomposed as $W \sim YY^T$, where $\mathrm{vec}(Y) \sim N(0, \Sigma \otimes I_n)$. Since $Y \in \mathcal{R}^{p\times n}$ has rank n with probability 1, the usual Moore-Penrose generalized inverse can be decomposed as

$$W^- \sim Y(Y^T Y)^{-2} Y^T \sim \Sigma^{1/2} Z (Z^T \Sigma Z)^{-2} Z^T \Sigma^{1/2}, \tag{3.1}$$

where $Z \in \mathcal{R}^{p\times n}$ is a matrix of iid standard normal variates. Write the spectral decomposition of Σ as $\Sigma = \Gamma\Lambda\Gamma^T$, where $\Gamma \in \mathcal{R}^{p\times p}$ is orthogonal and $\Lambda > 0$ is diagonal. Since the distribution of Z is invariant under orthogonal transformations we have $W^- \sim \Gamma\Lambda^{1/2} Z (Z^T \Lambda Z)^{-2} Z^T \Lambda^{1/2} \Gamma^T$. Consequently, without loss of generality, we assume that $\Sigma = \Lambda$ is a diagonal matrix when studying moments and other quantities. The left and right orthogonal transformations Γ and Γ^T can be restored straightforwardly for a general $\Sigma > 0$.

Let $M(\Lambda) = E(W^-)$. Then it follows that for all orthogonal matrices $P \in \mathcal{R}^{p\times p}$, $M(P^T \Lambda P) = P^T M(\Lambda) P$ and thus $M(\Lambda)$ is a tensor-valued isotonic tensor function of Λ. Isotropic tensor functions have been studied extensively in the literature on continuum mechanics (see [12] for an introduction and [9] for recent results). For instance, it is known from this literature that $M(\Lambda)$ and Λ have the same eigenvectors. Although there are various representations for isotropic tensor functions [9], they do not seem to provide further illumination in this setting. Let $V(\Lambda) = \mathrm{var}\{\mathrm{vec}(W^-)\} \in \mathcal{R}^{p^2 \times P^2}$. The variance $V(P^T \Lambda P) = (P^T \otimes P^T)V(\Lambda)(P \otimes P)$ is similarly structured as a fourth-order tensor function, and much the same comments apply.

When $\Lambda = I_p$, the distribution of W^- is invariant under orthogonal transformations, $W^- \sim PW^-P^T$ for all orthogonal $P \in \mathcal{R}^{p \times p}$. This invariance property was used extensively in the moment derivations for Theorem 2.1. However, when $\Lambda \neq I_p$ with distribution of W^- is no longer invariant and the moments of W^- become more complicated. Nevertheless, it is still possible to make some progress using symmetry arguments involving the rows z_i^T of Z, essentially utilizing invariance under a restricted class of transformations. This leads to the results stated in Theorem 3.1. In preparation, let $m_{ij}(\Lambda) = E\{z_i^T(Z^T\Lambda Z)^{-2})z_j\}$, let $v_{ij,kl}(\Lambda) = \mathrm{cov}\{z_i^T(Z^T\Lambda Z)^{-2}z_j, z_k^T(Z^T\Lambda Z)^{-2})z_l\}$, and let Λ_i denote the ith diagonal element of Λ, $i, j, k, l = 1, \ldots, p$.

Theorem 3.1. *Assume that $\Sigma = \Lambda$ is a diagonal matrix with diagonal elements Λ_i, $i = 1, \ldots, p$. Then $M(\Lambda)$ is a diagonal matrix with diagonal elements $M_{ii}(\Lambda) = \Lambda_i m_{ii}(\Lambda)$ and*

$$\begin{aligned}
V(\Lambda) \;=\; & \sum_{i,j=1}^{p} \Lambda_i\Lambda_j v_{ii,jj}(e_ie_j^T \otimes e_ie_j^T) + \sum_{i,j=1}^{p} \Lambda_i\Lambda_j v_{ij,ij}(e_je_j^T \otimes e_ie_i^T)(I_p + C_{p^2}) \\
& - 2\sum_{i=1}^{p} \Lambda_i^2 v_{ii,ii}(e_ie_i^T \otimes e_ie_i^T),
\end{aligned}$$

where the Λ arguments for $v_{(\cdot)}$ on the right hand side have been suppressed to improve readability, $v_{ii,jj} = \mathrm{cov}\{z_i^T(Z^T\Lambda Z)^{-2}z_i, z_j^T(Z^T\Lambda Z)^{-2})z_j\}$, $v_{ij,ij} = \mathrm{var}\{z_i^T(Z^T\Lambda Z)^{-2})z_j\}$ and $v_{ii,ii} = \mathrm{var}\{z_i^T(Z^T\Lambda Z)^{-2})z_i\}$.

The moments – m_{ii}, $v_{ij,ij}$, $v_{ii,jj}$ and $v_{ii,ii}$ – needed for Theorem 3.1 evidently do not have tractable closed-form representations.

3.2. Order of $\|\Sigma^{1/2}W_p^-(\Sigma, n)\Sigma^{1/2}\|$

Let $Z_0 = Z(Z^TZ)^{-1/2}$, and let $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and smallest eigenvalues of Σ. Then

$$\begin{aligned}
\Gamma^T W^- \Gamma &= \Lambda^{1/2}Z(Z^T\Lambda Z)^{-2}Z^T\Lambda^{1/2} \\
&= \Lambda^{1/2}Z(Z^TZ)^{-1}(Z_0^T\Lambda Z_0)^{-2}(Z^TZ)^{-1}Z^T\Lambda^{1/2}.
\end{aligned}$$

Because the normalized matrix Z_0 has orthogonal columns, $(Z_0^T\Lambda Z_0)^{-2} \leq \lambda_{\min}^{-2}I_p$ and thus $\Sigma^{1/2}W^-\Sigma^{1/2} \leq \lambda_{\min}^{-2}\Gamma\Lambda W_I^-\Lambda\Gamma^T$, where $W_I^- \sim W_p^-(I_p, n)$. The order of $\|\Sigma^{1/2}W^-\Sigma^{1/2}\|$ can now be found by application of Chebyschev's inequality.

Let $\epsilon > 0$ and, for notational convenience, let $H = \Sigma^{1/2}W^-\Sigma^{1/2}$. Then for all $h \in \mathcal{R}^p$ with $\|h\| = 1$,

$$\begin{aligned}
\Pr(h^T H h \geq \epsilon) &\leq \Pr(\lambda_{\min}^{-2} h^T \Gamma\Lambda W_I^- \Lambda\Gamma^T h \geq \epsilon) \\
&\leq \epsilon^{-2}\lambda_{\min}^{-4}\{\mathrm{var}(h^T\Gamma\Lambda W_I^-\Lambda\Gamma^T h) + E^2(h^T\Gamma\Lambda W_I^-\Lambda\Gamma^T h)\} \\
&= \epsilon^{-2}\lambda_{\min}^{-4}(h^T\Gamma\Lambda \otimes h^T\Gamma\Lambda)\mathrm{var}\{\mathrm{vec}(W_I^-)\}(\Lambda\Gamma^T h \otimes \Lambda\Gamma^T h) \\
&\quad + \epsilon^{-2}\lambda_{\min}^{-4} a_2^2 (h^T\Gamma\Lambda^2\Gamma^T h)^2 \\
&\leq \epsilon^{-2}(\lambda_{\max}/\lambda_{\min})^4\{2(c_2 + pd_2) + a_2^2\},
\end{aligned}$$

where $2(c_2 + pd_2) = \|\mathrm{var}\{\mathrm{vec}(W_I^-)\}\|$ as given in Corollary 2.2 and a_2 is as defined in Theorem 2.1. Combining this with Theorem 2.1 and the conclusions of Corollary 2.1 we have

Corollary 3.1. *Let $W \sim W_p(\Sigma, n)$.*

(i) Assume that $n > p + 3$ and that $p/n \to r$ with $0 \leq r < 1$. Then $\|\Sigma^{1/2}W^-\Sigma^{1/2}\| = O_p(n^{-1})$.

(ii) Assume that the condition number $\lambda_{\max}/\lambda_{\min}$ of Σ is bounded as $p \to \infty$, that $p > n + 3$ and that $n/p \to r$ with $0 < r < 1$. Then $\|\Sigma^{1/2}W^-\Sigma^{1/2}\| = O_p(n^{-1})$.

4. Proof of Theorem 2.1

The general idea of this proof is to use invariance arguments along with moment matching via Proposition 2.1.

A singular Wishart matrix $W \sim W_p(I_p, n)$, $p > n + 3$, can be decomposed as $W = YY^T$ with $\mathrm{vec}(Y) \sim N(0, I_p \otimes I_n)$ and $Y = H^T D^{1/2} U$, where $H^T \in \mathcal{R}^{p\times n}$ is semi-orthogonal $HH^T = I_n$, $U \in \mathcal{R}^{n\times n}$ is orthogonal and the diagonal elements $d_1, \ldots, d_n$ of the diagonal matrix $D \in \mathcal{R}^{n\times n}$ are non-zero with probability 1. Consequently, $W = H^T D H$ and $W^- = H^T D^{-1} H$. Moreover $Y^T Y \sim W_n(I_n, p)$ and since $p > n+3$, $Y^T Y$ has full rank with probability 1 and $(Y^T Y)^{-1} = U^T D^{-1} U \sim W_n^{-1}(I_n, p)$. Interchanging n and p in Proposition 2.1 we have

$$\begin{aligned}
E\{(Y^TY)^{-1}\} &= E(U^T D^{-1} U) = (p - n - 1)^{-1} I_n, && (4.1) \\
E\{(Y^TY)^{-1}(Y^TY)^{-1}\} &= E(U^T D^{-2} U) = c_3(p-1) I_n, && (4.2) \\
\mathrm{var}\{\mathrm{vec}(Y^TY)^{-1}\} &= \mathrm{var}\{\mathrm{vec}(U^T D^{-1} U)\} && (4.3) \\
&= c_3(I_{n^2} + C_{n^2}) + 2d_3 \mathrm{vec}(I_n)\mathrm{vec}^T(I_n),
\end{aligned}$$

where $c_3^{-1} = (p-n)(p-n-1)(p-n-3)$, $d_3^{-1} = (p-n)(p-n-1)^2(p-n-3)$.

4.1. $E(W^-)$

Using the fact that $Y \sim PY$ for any orthogonal matrix $P \in \mathcal{R}^{p\times p}$, we get

$$E(W^-) = E[\{(PY)(PY)^T\}^-] = PE(W^-)P^T$$

and consequently $E(W^-) = aI_p$ (see, for example, Eaton [5], Proposition 2.14). It remains to find a. Since $W^- = H^T D^{-1} H$,

$$ap = \mathrm{tr}\{E(W^-)\} = \mathrm{tr}\{E(H^T D^{-1} H)\} = \mathrm{tr}\{E(D^{-1})\} = n(p-n-1)^{-1},$$

where we used (4.1) for the last equality. From this we get $E(W^-)$; that is, $a = a_2$.

4.2. $\boldsymbol{E(W^- W^-)}$

Since $E(W^- W^-) = PE(W^- W^-)P^T$ for any orthogonal matrix P, $E(W^- W^-) = bI_p$. To find b we have

$$bp = \mathrm{tr}\{E(W^- W^-)\} = \mathrm{tr}\{E(H^T D^{-2} H)\} = \mathrm{tr}\{E(D^{-2})\} = c_3 n(p-1),$$

where we used (4.2) for the last equality. From this we conclude that $b = b_2$.

4.3. $\mathbf{var}\{\mathbf{vec}(\boldsymbol{W^-})\}$ *and* $\mathbf{var}\{\mathbf{tr}(\boldsymbol{W^-})\}$

Our proof of this part is based on the following proposition which gives a characterization of matrices that are invariant under a subclass of the orthogonal transformations.

Proposition 4.1. *Let $A \in \mathcal{R}^{p^2 \times p^2}$ such that $(P^T \otimes P^T)A(P \otimes P) = A$ for all orthogonal matrices $P \in \mathcal{R}^{p \times p}$. Then $A = cI_{p^2} + fC_{p^2} + 2d vec(I_p) vec^T(I_p)$, for some real multipliers c, d and f.*

This proposition is apparently well-known in the literature on continuum mechanics, where it is often referred to as a representation theorem for fourth-order isotropic tensors. Its proof for the case $p = 3$ can be found in the classical literature on Cartesian tensors [6]. Jog [7] provides a concise proof and cites ten other demonstrations of the same result, most of which are for $p = 3$. All of these proofs rely heavily on analytic traditions, notation and tensor operators that are not readily found in the statistical literature and might seem elusive on first reading. (A dictionary connecting tensors and common matrix operations in statistics was given by Dauxois et al. [3].) For completeness we have included in Section 4.5 a proof that does not use the technical machinery of continuum mechanics, but relies only on the Kronecker product, vec operator [8] and the commutation matrix [10, 11]. These operators were defined in the Introduction and are used widely in the statistical literature.

The collection of matrices $\mathcal{A}$ that satisfies the hypothesis of Proposition 4.1 forms a vector space over the real field that is closed under transposition and multiplication. The proposition essentially gives a basis $\{I_{p^2}, C_{p^2}, \mathrm{vec}(I_p)\mathrm{vec}^T(I_p)\}$ for $\mathcal{A}$.

For notational convenience, let $V = \mathrm{var}\{\mathrm{vec}(W^-)\}$. Since $Y \sim PY$ for any orthogonal matrix $P \in \mathcal{R}^{p \times p}$ we have

$$V = \mathrm{var}[\mathrm{vec}\{(PY)(PY)^T\}^-] = (P \otimes P)V(P^T \otimes P^T),$$

and consequently V satisfies the hypothesis of Proposition 4.1. However, V is also invariant under multiplication by C_{p^2}. Using Proposition 4.1 this implies that $c = f$: $V - VC_{p^2} = (c-f)I_{p^2} - (c-f)C_{p^2} = 0$. Consequently the class of covariance matrices must be of the form

$$\mathrm{var}\{\mathrm{vec}(W^-)\} = c(I_{p^2} + C_{p^2}) + 2d\mathrm{vec}(I_p)\mathrm{vec}^T(I_p). \tag{4.4}$$

It remains to find c and d which we do by moment matching.

From (4.4) we have

$$\mathrm{var}\{\mathrm{tr}(W^-)\} = \mathrm{var}\{\mathrm{tr}(D^{-1})\} = \mathrm{var}\{\mathrm{vec}^T(I_p)\mathrm{vec}(W^-)\} = 2cp + 2dp^2. \tag{4.5}$$

Now, by (4.3) we get another expression for

$$\mathrm{var}\{\mathrm{tr}(D^{-1})\} = \mathrm{var}\{\mathrm{vec}^T(I_n)\mathrm{vec}(Y^TY)^{-1}\} = 2c_3n + 2d_3n^2 \tag{4.6}$$

and using (4.5) and (4.6) together we see that c and d must satisfy

$$cp + dp^2 = c_3n + d_3n^2. \tag{4.7}$$

Since we are pursuing two factors – c and d – we require a second independent equation to determine them uniquely. This can be obtained by first taking the trace of (4.4) to get $\mathrm{tr}[\mathrm{var}\{\mathrm{vec}(W^-)\}] = cp^2 + cp + 2dp$. Second, we obtain a known expression for $\mathrm{tr}[\mathrm{var}\{\mathrm{vec}(W)^-\}]$ by writing it as

$$\mathrm{tr}[\mathrm{var}\{\mathrm{vec}(W^-)\}] = \mathrm{tr}[E\{\mathrm{vec}(W^-)\mathrm{vec}^T(W^-)\}] - \mathrm{tr}[E\{\mathrm{vec}(W^-)\}E^T\{\mathrm{vec}(W^-)\}]$$

and using previous results to reduce the right hand side. Using (4.2) and the previously derived form for $E(W^-)$ we have

$$\begin{aligned}
\mathrm{tr}[E\{\mathrm{vec}(W^-)\mathrm{vec}^T(W^-)\}] &= \mathrm{tr}[E\{\mathrm{vec}(H^TD^{-1}H)\mathrm{vec}^T(H^TD^{-1}H)\}] \\
&= \mathrm{tr}\{E(D^{-2})\} = nc_3(p-1), \\
\mathrm{tr}[E\{\mathrm{vec}(W^-)\}E^T\{\mathrm{vec}(W^-)\}] &= a_2^2p.
\end{aligned}$$

Consequently,

$$cp^2 + cp + 2dp = nc_3(p-1) - a_2^2p. \tag{4.8}$$

Using Maple to solve (4.7) and (4.8) for c and d gives the solutions stated in Theorem 2.1.

4.4. Moments of higher order

Moments of higher order can in principle be found similarly, by using results from von Rosen [17] in combination with moment matching. For instance, consider $E\{(W^-)^3\} = E(H^TD^{-3}H) = b_3I_p$. Proceeding as in Section 4.2, $E\{(Y^TY)^{-3}\} = E(U^TD^{-3}U) = r_3I_n$, where r_3 can be obtained from von Rosen's Corollary 3.1 again interchanging the roles of n and p. Consequently, $b_3 = nr_3/p$.

4.5. Proof of Proposition 4.1

The proof of Proposition 4.1 is based on using various subclasses of the orthogonal matrices to characterize the columns of A. This characterization is described in the next lemma, which uses the same hypothesis as the proposition.

Lemma 4.1. *Let $A \in \mathcal{R}^{p^2 \times p^2}$ be such that $(P^T \otimes P^T)A(P \otimes P) = A$ for all orthogonal matrices $P \in \mathcal{R}^{p \times p}$. Then for some real factors h, d, s_1 and s_2,*

(i) $A(e_i \otimes e_i) = hvec(I_p) + d(e_i \otimes e_i)$ *for* $i = 1, \ldots, p$.
(ii) $A(e_i \otimes e_j) = s_1(e_i \otimes e_j) + s_2(e_j \otimes e_i)$ *for* $i \neq j = 1, \ldots, p$.

Proof. This proof is based on taking various forms for P in the hypothesized relationship. We do not distinguish these forms notationally.

Part (i): Let P be any orthogonal matrix with the property that $Pe_i = e_i$ for a selected index i. Restricting consideration to this subclass of orthogonal matrices and multiplying the hypothesized equation on the right by $e_i \otimes e_i$ we have $(P^T \otimes P^T)A(e_i \otimes e_i) = A(e_i \otimes e_i)$. Let $M = \text{vec}^{-1}\{A(e_i \otimes e_i)\}$ so that

$$\text{vec}(P^T M P) = \text{vec}(M). \tag{4.9}$$

Without lost of generality take $i = p$, and consider orthogonal matrices of the form

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & 1 \end{pmatrix},$$

where $P_1 \in \mathcal{R}^{(p-1)\times(p-1)}$ is any orthogonal matrix. Clearly, $Pe_p = e_p$. Partition $M = (M_{jk})$, $j, k = 1, 2$ according to the partition of P. We consider the four partition components M_{jk} separately.

(1). From (4.9) we get $P_1^T M_{11} P_1 = M_{11}$ for all orthogonal matrices $P_1 \in \mathcal{R}^{p-1}$. It is well known that this implies $M_{11} = h_p I_{p-1}$ for some real multiplier h_p (see, for example, Eaton [5], Proposition 2.14).

(2). $P_1^T M_{12} = M_{12}$ for all orthogonal matrices P_1 implies that $M_{12} = 0$: Write $M_{12} = \lambda U$ with $U \in \mathcal{R}^{p-1}$ a semi-orthogonal matrix. Taking P_1^T to be an orthogonal matrix with row j equal to U^T, we have $M_{12} = \lambda e_j$ for any $j = 1 \ldots p$ and therefore $M_{12} = 0$. Similarly, $M_{21} = 0$.

(3). Finally, $M_{22} \in \mathcal{R}^1$ is arbitrary and it follows that M has the form $M = h_p I_p + d_p e_p e_p^T$ with $d_p = M_{22} - h_p$. Therefore, for $i = 1, \ldots, p$,

$$\text{vec}(M) = A(e_i \otimes e_i) = h_i \text{vec}(I_p) + d_i(e_i \otimes e_i). \tag{4.10}$$

It remains to show that h_i and d_i are constant over the index i. Take a new P such that $Pe_i = e_j$ and $Pe_j = e_i$ for two selected indices $i \neq j$. Then using (4.10), the hypothesis and (4.10) again we find

$$\begin{aligned} h_i \text{vec}(I_p) + d_i(e_i \otimes e_i) &= A(e_i \otimes e_i) \\ &= (P^T \otimes P^T)A(P \otimes P)(e_i \otimes e_i) \\ &= (P^T \otimes P^T)A(e_j \otimes e_j) \end{aligned}$$

$$\begin{aligned} &= \left(P^T \otimes P^T\right)\left(h_j \mathrm{vec}(I_p) + d_j(e_j \otimes e_j)\right) \\ &= h_j \mathrm{vec}(I_p) + d_j(e_i \otimes e_i). \end{aligned}$$

Therefore $h_i = h_j = h$, $d_i = d_j = d$ and part (i) follows.

Part (ii): The proof of this conclusion follows the same logic as the proof of part (i), but we use the subclass of orthogonal matrices with the property that $Pe_i = e_i$ and $Pe_j = e_j$ for selected indices $i \neq j$. Restricting consideration to this subclass and multiplying the hypothesized equation on the right by $e_i \otimes e_j$ we have $(P^T \otimes P^T)A(e_i \otimes e_j) = A(e_i \otimes e_j)$. Let $M = \mathrm{vec}^{-1}\{A(e_i \otimes e_j)\}$ so that $\mathrm{vec}(P^T MP) = \mathrm{vec}(M)$. Without lost of generality take $i = p-1$ and $j = p$, and consider orthogonal matrices of the form

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & I_2 \end{pmatrix},$$

where $P_1 \in \mathcal{R}^{(p-2)\times(p-2)}$ is any orthogonal matrix. Partition $M = (M_{jk})$, $j, k = 1, 2$ to conform with the partitions of P. Using the hypothesis of the lemma, we reason as follows. (1) $P_1^T M_{11} P_1 = M_{11}$ for all P_1 orthogonal matrices of order $p-2$ again implies that $M_{11} = c_{ij} I_{p-2}$ for some real factor c_{ij}. (2) $P_1^T M_{12} = M_{12}$ for all orthogonal matrices P_1 implies again $M_{12} = 0$, and analogously $M_{21} = 0$. (3) M_{22} is arbitrary and therefore

$$A(e_i \otimes e_j) = c_{ij}\mathrm{vec}(I_p) + s_{1ij}(e_i \otimes e_j) + s_{2ij}(e_j \otimes e_i) + t_{ij}(e_i \otimes e_i) + u_{ij}(e_j \otimes e_j). \tag{4.11}$$

It follows from part (i) and the fact that A^T also satisfies the hypothesis that $(e_k^T \otimes e_k^T)A(e_i \otimes e_j) = 0$ for $i \neq j = 1, \ldots, p$ and $k = 1, \ldots, p$. Using this result and multiplying (4.11) by $(e_k^T \otimes e_k^T)$, with $k \neq i$ and $k \neq j$, $(e_i^T \otimes e_i^T)$ and $(e_j^T \otimes e_j^T)$ respectively we conclude that $c_{ij} = t_{ij} = u_{ij} = 0$ for $i \neq j$. Therefore

$$A(e_i \otimes e_j) = s_{1ij}(e_i \otimes e_j) + s_{2ij}(e_j \otimes e_i),\ i \neq j. \tag{4.12}$$

It remains to show that s_{1ij} and s_{2ij} are constant in the indices $i \neq j$. Take a new subclass of P's such that, for two additional selected indices $k \neq s$, $Pe_i = e_k$, and $Pe_j = e_s$ where still $i \neq j$. Then using (4.12), the hypothesis and (4.12) again we get that

$$\begin{aligned} s_{1ij}(e_i \otimes e_j) + s_{2ij}(e_j \otimes e_i) &= (P^T \otimes P^T)A(P \otimes P)(e_i \otimes e_j) \\ &= (P^T \otimes P^T)A(e_k \otimes e_s) \\ &= \left(P^T \otimes P^T\right)\left(s_{1ks}(e_k \otimes e_s) + s_{2ks}(e_s \otimes e_k)\right) \\ &= s_{1ks}(e_i \otimes e_j) + s_{2ks}(e_j \otimes e_i). \end{aligned}$$

Multiplying the first and the last term by $(e_i^T \otimes e_j^T)$ and $(e_j^T \otimes e_i^T)$ respectively we get

$$A(e_i \otimes e_j) = s_1(e_i \otimes e_j) + s_2(e_j \otimes e_i), \quad \text{for } i \neq j,$$

which concludes the proof of the lemma. $\square$

Turning to the proof of Proposition 4.1, we first show that the multipliers in Lemma 4.1 are functionally related; in particular, $d = s_1 + s_2$. It follows immediately from part (ii) of Lemma 4.1 that, for $i \neq j$, $A(P \otimes P)(e_i \otimes e_j) = (P \otimes P)\,(s_1(e_i \otimes e_j) + s_2(e_j \otimes e_i))$. Taking P to be in the subclass of orthogonal matrices with the property that, for selected indices $i \neq j$, $Pe_i = (e_i + e_j)/\sqrt{2}$ and $Pe_j = (e_j - e_i)/\sqrt{2}$, we have immediately that

$$A\left(\frac{e_i + e_j}{\sqrt{2}} \otimes \frac{e_j - e_i}{\sqrt{2}}\right) \quad = \quad s_1\left(\frac{e_i + e_j}{\sqrt{2}} \otimes \frac{e_j - e_i}{\sqrt{2}}\right) + s_2\left(\frac{e_j - e_i}{\sqrt{2}} \otimes \frac{e_i + e_j}{\sqrt{2}}\right)$$

Expanding this equation and simplifying we find that

$$d((e_j \otimes e_j) - (e_i \otimes e_i)) \quad = \quad (s_1 + s_2)\,((e_j \otimes e_j) - (e_i \otimes e_i))$$

and consequently $d = s_1 + s_2$. The conclusion of the proposition now follows from Lemma 4.1 with $d = s_1 + s_2$: Taking $v = \sum_{i,j} c_{ij}(e_i \otimes e_j)$,

$$\begin{aligned} Av &= A\left(\sum_{i,j} c_{ij}(e_i \otimes e_j)\right) \\ &= \sum_{i \neq j} c_{ij}\,(s_1(e_i \otimes e_j) + s_2(e_j \otimes e_i)) + \sum_{i=1}^{p} c_{ii}\,(h\mathrm{vec}(I_p) + (s_1 + s_2)(e_i \otimes e_i)) \\ &= s_1 \sum_{i,j} c_{ij}(e_i \otimes e_j) + s_2 \sum_{i,j} c_{ij}(e_j \otimes e_i) + h \sum_{i=1}^{p} c_{ii}\mathrm{vec}(I_p) \\ &= s_1 v + s_2 C_{p^2} v + h_1 \mathrm{vec}(I_p)\mathrm{vec}^T(I_p) v, \end{aligned}$$

where $h_1 = h \sum_{i=1}^{p} c_{ii}$.

5. Proof of Theorem 3.1

Recall that Theorem 3.1 requires $\Sigma = \Lambda$ to be a diagonal matrix. For notational convenience, let $H = (Z^T \Lambda Z)^{-2}$. The conclusion that $M(\Lambda)$ is a diagonal matrix arises by noting that, for $i \neq j$, $z_i^T H z_j \sim -z_i^T H z_j$, and thus $E(z_i^T H z_j) = 0$.

By a similar symmetry argument, the element $v_{ij,kl}$ of V equals 0 when at least one of its indices i, j, k, l is distinct; that is, not equal to any other index. If no indices are distinct then they must be equal in pairs, leading to four possibilities: for $i \neq j$, $v_{ii,jj}$, $v_{ij,ij}$, $v_{ij,ji}$ and, for $i = j$, $v_{ii,ii}$. However, $v_{ij,ij} = v_{ij,ji}$, which leads to the three v terms in the Theorem. The form of V follows from these results, the representation $V = \sum_{ij,kl} (\Lambda_i \Lambda_k \Lambda_k \Lambda_l)^{1/2} v_{ij,kl} \times \mathrm{vec}(e_i e_j^T)\mathrm{vec}(e_k e_l^T)$, and the definition of the commutation matrix.

References

[1] Basser, P. J. and Pajevic, S. (2007). Spectral decomposition of a 4th order covariance tensor: Applications to diffusion tensor MRI. *Signal Processing* **87**, 220–236.

[2] Bodnar, T. and Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *J. Multivariate Anal.* **99**, 2389–2405. MR2463397

[3] Dauxois, J., Romain, Y. and Viguier-Pla, S. (1994). Tensor products and statistics. *Linear Algebra and its Applications* **210**, 59–88. MR1294771

[4] Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2006). Distribution of the generalized inverse of a random matrix and its applications. *Journal of Statistical Planning and Inference* **136**, 183–192. MR2207179

[5] Eaton, M. L. (2007). *Multivariate Statistics: A Vector Space Approach.* Beachwood, Ohio: Institute of Mathematical Statistics. MR2431769

[6] Jeffreys, H. (1931). *Cartesian Tensors.* Cambridge: Cambridge University Press. Page 91. MR0133075

[7] Jog, C. S. (2006). A concise proof of the representation theorem for fourth-order isotropic tensors. *J. Elasticity* **85**, 119–124. MR2265724

[8] Henderson, H, V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canadian J. Statist.* **7**, 65–81. MR0549795

[9] Itskov, M. (2009). *Tensor Analysis and Tensor Algebra for Engineers, 2nd Edition.* New York: Springer.

[10] Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* **7**, 381–394. MR0520247

[11] Neudecker, H. and Wansbeek, T. (1983). Some Results on Commutation Matrices, with Statistical Applications. *Canadian J. Statist.* **11**, 221–231. MR0732996

[12] Ogden, R. W. (2001). Elements of the theory of finite elasticity. In Fu, Y. B. and Ogden, R. W. (eds) *Nonlinear Elasticity: Theory and Applications.* Cambridge University Press, 1–58. MR1835108

[13] Stone, M. (1987). *Coordinate-Free Multivariate Statistics.* New York: Oxford University Press. MR0916472

[14] Srivastava, M. S. (2003). Singular Wishart and multivariate beta distributions. *Ann. Statist.* **31**, 1537–1560. MR2012825

[15] Tyler, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9**, 725–736. MR0619278

[16] Uhlig, H. (1994). On singular Wishart and singular multivariate beta distributions. *Ann. Statist.* **22**, 395–405. MR1272090

[17] von Rosen, D. (1988). Moments for the inverted Wishart distribution. *Scand. J. Statist.* **15**, 97–109. MR0968156

[18] Zhang, Z. (2007). Pseudo-inverse multivariate/matrix-variate distributions. *J. Multivariate Anal.* **98**, 1684–1692. MR2370113