$$\begin{bmatrix} \text{I L} \\ \text{A S} \end{bmatrix}$$

ON THE MEAN AND DISPERSION OF THE MOORE-PENROSE GENERALIZED INVERSE OF A WISHART MATRIX[*]

SHINPEI IMORI[†] AND DIETRICH VON ROSEN[‡]

Abstract. The Moore-Penrose inverse of a singular Wishart matrix is studied. When the scale matrix equals the identity matrix the mean and dispersion matrices of the Moore-Penrose inverse are known. When the scale matrix has an arbitrary structure no exact results are available. The article complements the existing literature by deriving upper and lower bounds for the expectation and an upper bound for the dispersion of the Moore-Penrose inverse. The results show that the bounds become large when the number of rows (columns) of the Wishart matrix are close to the degrees of freedom of the distribution.

1. Introduction. In this article, all matrices are real valued. Let the matrix W: $p \times p$ be Wishart distributed with n degrees of freedom which will be denoted $W \sim W_p(\Sigma, n)$, where Σ can be considered to be a positive definite dispersions matrix. More precisely, there exists a matrix normally distributed random variable $X \sim N_{p,n}(0, \Sigma, I_n)$ such that $W = XX'$, where $N_{p,n}(\bullet, \bullet, \bullet)$ denotes the matrix normal distribution with the dispersion $D[X] = I_n \otimes \Sigma$, $'$ denotes the transpose and $\otimes$ denotes the Kronecker product.

Throughout this note it will be assumed that $p > n$ which can be motivated from a high-dimensional perspective when there are p dependent variables which distribution depends on "many" parameters, in our case the unstructured Σ, and less independent observations n. Since under this condition W is singular, we will be interested in the Moore-Penrose inverse of W, which is written W^+.

In statistics when W^{-1} exists one often uses functions of W^{-1}. For example, in discriminant analysis the linear discriminant function for $y_i \sim N_p(\mu_1, \Sigma)$, $i \in \{1, \dots, n_1\}$, and $z_j \sim N_p(\mu_2, \Sigma)$, $j \in \{1, \dots, n_2\}$, if μ_1, μ_2 and Σ are known and x is an observation which is to be classified, is based on

$$(1.1) \qquad D(x; \mu_1, \mu_2, \Sigma^{-1}) = (\mu_1 - \mu_2)' \Sigma^{-1}(x - (\mu_1 + \mu_2)/2).$$

Put $n = n_1 + n_2$. If $n > (p+1)$ and the parameters μ_1, μ_2 and Σ are unknown they can be replaced by their maximum likelihood estimators, in particular Σ^{-1} is replaced by nW^{-1}, where the sums of squares matrix W satisfies $W \sim W_p(\Sigma, n - 2)$, which yields the classification function $D(x; \widehat{\mu}_1, \widehat{\mu}_2, W^{-1})$ with $\widehat{\mu}_i$ denoting the maximum likelihood estimator of μ_i, $i \in \{1, 2\}$. Another example is the weighted least squares estimator (maximum likelihood estimator) for the Growth Curve model, i.e., $Y \sim N_{p,n}(ABC, \Sigma, I_n)$, where A: $p \times q$, $q < p$ and C: $k \times n$ are known matrices, and $\{B, \Sigma\}$ are unknown parameter matrices (see [6,7]). Under full

[†]Graduate School of Science, Hiroshima University, Japan (imori@hiroshima-u.ac.jp).
[‡]Energy and Technology, Swedish University of Agricultural Sciences, Uppsala, Sweden, and Department of Mathematics, Linköping University, Sweden (dietrich.von.rosen@slu.se).

$$\boxed{\begin{array}{c} \text{I\ L} \\ \text{A\ S} \end{array}}$$

rank conditions and $p \le n - k$ the maximum likelihood estimator of B equals

$$(1.2) \qquad \widehat{B} = (A'W^{-1}A)^{-1}A'W^{-1}YC'(CC')^{-1},$$

where $W = Y(I_n - C'(CC')^{-1}C)Y' \sim W_p(\Sigma, n - k)$. When $p > (n - 2)$ in the discriminant function or $p > (n - k)$ in the Growth Curve model one sometimes replaces W^{-1} by W^+ since W^{-1} does not exist (see, e.g., [4,9]). Thus, when p is "larger" than n, instead of (1.1) and (1.2)

$$D(x; \widehat{\mu}_1, \widehat{\mu}_2, W^+),$$
$$(1.3) \qquad \widehat{B} = (A'W^+A)^{-1}A'W^+XC'(CC')^{-1}$$

are used. Of course, (1.3) is no longer a maximum likelihood estimator and some more conditions are needed so that $(A'W^+A)^{-1}$ exists. To replace W^{-1} by W^+, when p is "larger" than n, is, however, often unclear why this can take place.

If $W \sim W_p(\Sigma, n)$ then as noted before $W = XX'$ for some $X \sim N_{p,n}(0, \Sigma, I_n)$ and if $p > n$

$$(1.4) \qquad W^+ = X(X'X)^{-1}(X'X)^{-1}X' = (X^+)'X^+.$$

This is a well known relation and follows from the four defining conditions of the Moore-Penrose inverse:

$$WW^+W = W, \quad W^+WW^+ = W^+,$$

$$(WW^+)' = WW^+, \quad (W^+W)' = W^+W.$$

If $n \ge p$, W^+ reduces to W^{-1}. Moreover, the density for W^{-1} (when $n \ge p$) is well known and it follows directly from the Wishart density by the transformation $W \to W^{-1}$ and using the Jacobian $|W|^{-1/2(p+1)}$.

Concerning the density of a Moore-Penrose inverse there are some results available when a matrix Z, has full column rank. In this case, $Z^+ = (Z'Z)^{-1}Z'$ has a density $|Z^+(Z^+)'|^{-p}f(Z)$, where $f(Z)$ is the density for Z (see [2,5,10,11]). From this expression, in principle, the density for W^+ can be found and advanced direct calculations of Jacobians leads to the density for W^+ (see [5,11]).

To derive moments via the density is however not easy. One reason for the difficulty is that the density expression is a function of the eigen values which is difficult to handle. Moreover, let A be a non-singular square matrix and then AW^+A' does not equal $((A')^{-1}WA^{-1})^+$ unless A is an orthogonal matrix. Contrary to, if W^{-1} exists, then $AW^{-1}A' = ((A')^{-1}WA^{-1})^{-1}$ which often is used in calculations. This implies that if considering W^+ it matters if Σ in $W_p(\Sigma, n)$ equals I_p or differs from the identity matrix. Moreover, Cook and Forzani [1] in an interesting article presented a number of results when $\Sigma = I_p$. They also discuss when Σ is an arbitrary positive definite matrix and find some approximations of the mean and dispersion matrix, i.e., $E[W^+]$ and $D[W^+]$. In this article, we complement their results, in particular, by deriving upper bounds of these moments.

We are interested in $E[W^+]$ and $D[W^+]$, and it seems for our purposes difficult to utilize the density for W^+. If $W \sim W_p(\Sigma, n)$ and $p + 1 < n$, then

$$(1.5) \qquad E[W^{-1}] = \frac{1}{n - p - 1}\Sigma^{-1},$$

and if $p + 3 < n$,

$$(1.6) \qquad D[W^{-1}] = c_1(I_{p^2} + K_{p,p})(\Sigma^{-1} \otimes \Sigma^{-1}) + \left\{ c_2 - \frac{1}{(n-p-1)^2} \right\} \text{vec}\Sigma^{-1}\text{vec}'\Sigma^{-1},$$

where vec stands for the *vec*-operator, $K_{p,p}$ is the commutation matrix (see, e.g., [7, Section 1.3]) and

$$c_1 = \frac{1}{(n-p)(n-p-1)(n-p-3)}, \quad c_2 = (n-p-2)c_1.$$

If $p > n+1$ and $\Sigma = I_p$ it can be shown that

$$(1.7) \qquad E[W^+] = \frac{n}{p(p-n-1)}I_p.$$

One way of proving this relation is, since for $\Sigma = I_p$ and for all orthogonal matrices Γ, $\Gamma W^+\Gamma'$ has the same distribution as W^+, it follows that $E[W^+] = cI_p$ for some positive constant c. The constant can be determined by taking the trace, i.e., $E[\text{tr}\{W^+\}] = cp$ and using (1.4):

$$\text{tr}\{W^+\} = \text{tr}\{X(X'X)^{-1}(X'X)^{-1}X'\} = \text{tr}\{(X'X)^{-1}\},$$

where now $X'X \sim W_n(I_n, p)$ which yields (1.7), since $E[\text{tr}\{(X'X)^{-1}\}] = n/(p-n-1)$, and thus, $c = n/\{p(n-p-1)\}$.

The first statement in the next theorem has hereby been verified. For the second statement it is referred to [1]. However, it can be noted that due to invariance with respect to orthogonal transformations it is enough to know $E[(\text{tr}\{W^+\})^2] = E[(\text{tr}\{(X'X)^{-1}\})^2]$ and $E[\text{tr}\{W^+W^+\}] = E[\text{tr}\{(X'X)^{-1}(X'X)^{-1}\}]$ which can be obtained from (1.6).

PROPOSITION 1.1. (Cook and Forzani [1]) *Let $W \sim W_p(I_p, n)$. Then,*

(i) *if $p > n+1$, $E[W^+] = a_1 I_p$, where $a_1 = n/\{p(p-n-1)\}$;*
(ii) *if $p > n+3$,*

$$E[\text{vec}W^+\text{vec}'W^+] = a_2(I_{p^2} + K_{p,p}) + a_3\text{vec}I_p\text{vec}I_p',$$

where

$$a_2 = \frac{n\{p(p-1) - n(p-n-2) - 2\}}{p(p-1)(p+2)(p-n)(p-n-1)(p-n-3)},$$

$$a_3 = \frac{n\{4 + n(p+1)(p-n-2)\}}{p(p-1)(p+2)(p-n)(p-n-1)(p-n-3)}.$$

2. Preparation. In this section, mainly some useful lemmas are presented. Let $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ be the ordered eigen values of a symmetric matrix A: $n \times n$. Moreover, $A \geq 0$ ($A > 0$) means that A is positive semi-definite (positive definite) and $A \geq B$ means that $A - B \geq 0$, where both A and B are supposed to be positive semi-definite. Concerning ordering of matrices the following definitions will be used.

DEFINITION 2.1. Let U and V be positive semi-definite matrices.

(i) (Löwner ordering) If, for all vectors α of proper size, $\alpha'U\alpha \leq \alpha'V\alpha$, then it is written $U \leq V$.

(ii) If, for all vectors α of proper size, $\alpha' E[U]\alpha \leq \alpha' E[V]\alpha$, then it is written $E[U] \leq E[V]$.

(iii) Let $U = U_1 \otimes U_2$ and $V = V_1 \otimes V_2$, where all matrices are supposed to be positive semi-definite. If, for all vectors α of proper size,

$$(\alpha \otimes \alpha)'U(\alpha \otimes \alpha) \leq (\alpha \otimes \alpha)'V(\alpha \otimes \alpha),$$

then it is written $U \preceq V$.

(iv) If, for all vectors α of proper size,

$$(\alpha \otimes \alpha)'D[U](\alpha \otimes \alpha) \leq (\alpha \otimes \alpha)'D[V](\alpha \otimes \alpha),$$

then it is written $D[U] \preceq D[V]$, i.e., $D[\alpha'U\alpha] \leq D[\alpha'V\alpha]$.

The condition in (i) is identical to the condition of positive semi-definiteness of $V - U$. Note that in (iii) $U_1 \otimes U_2$ also can be of the form $\mathrm{vec}U_1\mathrm{vec}'U_2$ or $K_{\bullet,\bullet}(U_1 \otimes U_2)$ since $\mathrm{vec}U_1\mathrm{vec}'U_2 = \mathrm{vec}U_1 \otimes \mathrm{vec}'U_2$ (forgetting that $\mathrm{vec}U_i$ cannot be positive semi-definite) and $(\alpha \otimes \alpha)'K_{\bullet,\bullet}(U_1 \otimes U_2)(\alpha \otimes \alpha) = (\alpha \otimes \alpha)'(U_1 \otimes U_2)(\alpha \otimes \alpha)$.

Some obvious but useful results are presented in the next lemma.

LEMMA 2.2.

(i) If $A_i \preceq B_i$, $i \in \{1, 2\}$, then $A_1 + A_2 \preceq B_1 + B_2$.

(ii) If $A_i \leq B_i$, $i \in \{1, 2\}$, then $A_1 \otimes A_2 \preceq B_1 \otimes B_2$.

Proof. Note that

$$(\alpha \otimes \alpha)'(A_1 + A_2)(\alpha \otimes \alpha)$$
$$= (\alpha \otimes \alpha)'A_1(\alpha \otimes \alpha) + (\alpha \otimes \alpha)'A_2(\alpha \otimes \alpha)$$
$$\leq (\alpha \otimes \alpha)'B_1(\alpha \otimes \alpha) + (\alpha \otimes \alpha)'B_2(\alpha \otimes \alpha)$$
$$= (\alpha \otimes \alpha)'(B_1 + B_2)(\alpha \otimes \alpha),$$

and (i) has been established. Moreover,

$$(\alpha \otimes \alpha)'(A_1 \otimes A_2)(\alpha \otimes \alpha)$$
$$= \alpha'A_1\alpha\alpha'A_2\alpha$$
$$\leq \alpha'B_1\alpha\alpha'B_2\alpha$$
$$= (\alpha \otimes \alpha)'(B_1 \otimes B_2)(\alpha \otimes \alpha),$$

and (ii) is verified.   □

The next lemma presents a well known result whereas in a third lemma a more specific result is given.

LEMMA 2.3. (Poincaré separation theorem) *Let L: $n \times p$ satisfy $LL' = I_n$ and let A: $p \times p$ be a symmetric matrix. Then, for $i \in \{1, \ldots, n\}$,*

(i) $\lambda_i(LAL') \leq \lambda_i(A)$;

(ii) $\lambda_i(LAL') \geq \lambda_{p-n+i}(A)$.

LEMMA 2.4. *Let L: $n \times p$ satisfy $LL' = I_n$ and let*

$$P_0 = L'(L\Sigma L')^{-1}(L\Sigma L')^{-1}L,$$

where $\Sigma > 0$. Then

(i) $(\lambda_p(\Sigma^{-1}))^2 L'L \le P_0 \le (\lambda_1(\Sigma^{-1}))^2 L'L$;

(ii) $P_0 \le \lambda_1(\Sigma^{-1})\Sigma^{-1}$.

Proof. The lower inequality of the statement (i) follows if a λ can be found such that

$$(2.8) \qquad (L\Sigma L')^{-1}(L\Sigma L')^{-1} - \lambda I_n \ge 0.$$

Thus, a value of λ has to be determined which is smaller or equal to

$$\lambda_n((L\Sigma L')^{-1}(L\Sigma L')^{-1}).$$

Lemma 2.3 (i) will be used and

$$\lambda_n((L\Sigma L')^{-1}(L\Sigma L')^{-1}) = (\lambda_1(L\Sigma L'L\Sigma L'))^{-1} = (\lambda_1(\Sigma L'L\Sigma))^{-1}$$
$$= (\lambda_1(L\Sigma\Sigma L'))^{-1} \ge (\lambda_1(\Sigma\Sigma))^{-1} = (\lambda_p(\Sigma^{-1}))^2,$$

and the lower inequality of (i) has been verified. The upper inequality of (i) can be proven in the same manner.

Concerning statement (ii)

$$\Sigma^{-1/2}(\lambda_1(\Sigma^{1/2}P_0\Sigma^{1/2})I_p - \Sigma^{1/2}P_0\Sigma^{1/2})\Sigma^{-1/2} \ge 0$$

and

$$\lambda_1(\Sigma^{1/2}P_0\Sigma^{1/2}) = \lambda_1((L\Sigma L')^{-1}) = (\lambda_n(L\Sigma L'))^{-1} \le (\lambda_p(\Sigma))^{-1},$$

where the inequality is based on Lemma 2.3 (ii). □

3. Main results. The aim is to determine bounds for $E[W^+]$ and $D[W^+]$, $W \sim W_p(\Sigma, n)$, in the case when $p > n$ and $\Sigma > 0$ is unstructured.

3.1. Upper and lower bounds for $E[W^+]$. When $W \sim W_p(\Sigma, n)$ there exists a normally distributed $X \sim N_{p,n}(0, \Sigma, I_n)$ such that $W = XX'$. Let $Y = \Sigma^{-1/2}X$ and then, due to (1.4),

$$(3.9) \qquad E[W^+] = (2\pi)^{-np/2} \int e^{-\mathrm{tr}\{YY'\}/2}\Sigma^{1/2}Y(Y'\Sigma Y)^{-1}(Y'\Sigma Y)^{-1}Y'\Sigma^{1/2}\, dY$$

will be studied. Now make the variable substitution $Y' = TL$, where $LL' = I_n$, L: $n \times p$, and $T = (t_{ij})$ is lower triangular with positive diagonal elements. The Jacobian of this transformation equals (see, e.g., [3, Theorem 1.4.20] and [8, p. 38])

$$(3.10) \qquad |J(Y \to T, L)|_+ = \prod_{i=1}^n t_{ii}^{p-i}g(L),$$

where $g(L) = \prod_{i=1}^n |L_i|_+$, $L_i = (\ell_{jk})$, $j, k \in \{1, \ldots, i\}$ and the functionally independent elements in L are $\ell_{12}, \ell_{13}, \ldots, \ell_{1p}, \ell_{23}, \ldots, \ell_{2p}, \ldots, \ell_{n1}, \ldots, \ell_{np}$. Here $|\bullet|_+$ denotes the absolute value of the determinant. Thus, instead of (3.9), one has

$$E[W^+] = (2\pi)^{-np/2} \int e^{-\mathrm{tr}\{T'T\}/2}\Sigma^{1/2}L'(L\Sigma L')^{-1}T^{-1}(T')^{-1}(L\Sigma L')^{-1}L\Sigma^{1/2}$$
$$(3.11) \qquad \times \prod_{i=1}^n t_{ii}^{p-i}g(L)\, dL\, dT.$$

Put $V = T'T$ and (see, e.g., [3, Theorem 1.4.18])

$$|J(T \to V)|_+ = 2^{-n} \prod_{j=1}^{n} t_{jj}^{-(n-j+1)}.$$

Then

$$E[W^+] = (2\pi)^{-np/2} 2^{-n} \int e^{-\mathrm{tr}\{V\}/2} \Sigma^{1/2} L' (L\Sigma L')^{-1} V^{-1} (L\Sigma L')^{-1} L\Sigma^{1/2}$$
$$\times |V|^{(p-n-1)/2} g(L) \, dL \, dV.$$

If $p - n - 1 > 0$ it follows from the expectation of the inverse Wishart matrix in (1.5) that

(3.12) $$E[W^+] = (p-n-1)^{-1} c(n,p)^{-1} \int \Sigma^{1/2} L' (L\Sigma L')^{-1} (L\Sigma L')^{-1} L\Sigma^{1/2} g(L) \, dL,$$

where $c(n,p) = (2\pi)^{np/2} 2^n s(n,p)$ and $s(n,p)$ is the standardization constant in a Wishart density for a $W_n(I_n, p)$-variable, i.e.,

$$s(n,p) \int e^{-\mathrm{tr}\{V\}/2} |V|^{(p-n-1)/2} \, dV = 1.$$

Before proceeding a lemma is presented which can be used to integrate out $g(L)$ from certain forthcoming expressions.

LEMMA 3.1. *Let $g(L)$ be as in* (3.10), $LL' = I_n$ *and $s(n,p)$ is as in* (3.12). *Then*

(i) $\int g(L) \, dL = c(n,p)$;
(ii) $\int L'Lg(L) \, dL = np^{-1} c(n,p) I_p$.
(iii) $\int (\alpha' L' L\alpha)^2 g(L) \, dL = (2n + n^2)(2p + p^2)^{-1} c(n,p)(\alpha'\alpha)^2$, *for all $\alpha \in \mathbb{R}^p$*.

Proof. Let $Y \sim N_{p,n}(0, I_p, I_n)$, $p > n$, and then

$$1 = (2\pi)^{-np/2} \int e^{-\mathrm{tr}\{YY'\}/2} dY.$$

Make the same variable transformations as in the beginning of this section, i.e., $Y' = TL$, $V = T'T$ and we end up with the expression

$$1 = c(n,p)^{-1} s(n,p) \int |V|^{(p-n-1)/2} e^{-\mathrm{tr}\{V\}/2} g(L) \, dV \, dL$$

which after integrating out V establishes (i), where we can assume that $V \sim W_n(I_n, p)$. To verify (ii) it is started with the known integral $E[YY'] = nI_p$, i.e.,

$$nI_p = (2\pi)^{-np/2} \int e^{-\mathrm{tr}\{YY'\}/2} YY' \, dY.$$

Once again making the variable transformations $Y' = TL$, $V = T'T$ yields

$$nI_p = c(n,p)^{-1} s(n,p) \int |V|^{(p-n-1)/2} e^{-\mathrm{tr}\{V\}/2} L'VLg(L) \, dV \, dL$$
$$= c(n,p)^{-1} p \int L'Lg(L) \, dL$$

and (ii) has been shown.

Finally, we show (iii). A moment relation for the Wishart distribution yields that

$$E[\text{vec}(YY')\text{vec}'(YY')] = n(I_{p^2} + K_{p,p}) + n^2\text{vec}I_p\text{vec}'I_p$$
$$= (2\pi)^{-np/2}\int e^{-\text{tr}\{YY'\}/2}\text{vec}(YY')\text{vec}'(YY')\,dY.$$

From the same arguments for proving (i) and (ii), it follows that

$$n(I_{p^2} + K_{p,p}) + n^2\text{vec}I_p\text{vec}'I_p = c(n,p)^{-1}s(n,p)\int |V|^{(p-n-1)/2}e^{-\text{tr}\{V\}/2}(L\otimes L)'\text{vec}V$$
$$\times\text{vec}'V(L\otimes L)\,g(L)\,dV\,dL$$
$$= c(n,p)^{-1}\int (L\otimes L)'\{p(I_{n^2} + K_{n,n}) + p^2\text{vec}I_n\text{vec}'I_n\}(L\otimes L)\,g(L)\,dL.$$

Note that

$$(\alpha\otimes\alpha)'(L\otimes L)'K_{n,n}(L\otimes L)(\alpha\otimes\alpha) = (\alpha'L'L\alpha)^2,$$
$$(\alpha\otimes\alpha)'(L\otimes L)'\text{vec}I_n = \alpha'L'L\alpha.$$

These relations imply that

$$(2n + n^2)(\alpha'\alpha)^2 = (2p + p^2)c(n,p)^{-1}\int (\alpha'L'L\alpha)^2\,g(L)\,dL. \qquad \square$$

Note that the proof of (i) follows a possible way of deriving the Wishart density (see, e.g., [8, Corollary 3.2.1]).

THEOREM 3.2. *Let $W \sim W_p(\Sigma, n)$, $p > n + 1$ and $\Sigma > 0$. Then, in the sense of Definition 2.1 (ii),*

$$a_1(\lambda_p(\Sigma^{-1}))^2\Sigma \leq E[W^+] \leq a_1(\lambda_1(\Sigma^{-1}))^2\Sigma,$$

where a_1 is given in Proposition 1.1 (i).

Proof. Put

(3.13) $$P = \Sigma^{1/2}L'(L\Sigma L')^{-1}(L\Sigma L')^{-1}L\Sigma^{1/2}.$$

It follows from Lemma 2.4 (i) that

$$(\lambda_p(\Sigma^{-1}))^2\Sigma^{1/2}L'L\Sigma^{1/2} \leq P \leq (\lambda_1(\Sigma^{-1}))^2\Sigma^{1/2}L'L\Sigma^{1/2}.$$

Using (3.12) and Lemma 3.1 (ii) yield the statement (i). $\square$

Note that the bounds presented in Theorem 3.2 are sharp in the sense that the upper and lower bounds, if $\Sigma = I_p$, are identical and equal the expectation in (1.7).

A consequence of the theorem is that if p is close to $n + 1$ the Moore-Penrose inverse nW^+ can be a poor estimator of Σ^{-1} because, in this case, a_1 becomes large. It is also noted that if a_1 becomes large this has nothing to do with the true Σ^{-1}. This means that in many high-dimensional problems the main problem occurs when p is close to n and not when p is much larger than n, for example when considering the estimator (1.3) of the mean parameter in the Growth Curve model.

Another upper bound is presented in the next theorem.

THEOREM 3.3. *Let $W \sim W_p(\Sigma, n)$, $p > n + 1$ and $\Sigma > 0$. Then, in the sense of Definition 2.1 (ii),*

$$E[W^+] \leq \frac{1}{p-n-1}\lambda_1(\Sigma^{-1})I_p.$$

Proof. Let P be as in Theorem 3.2. According to Lemma 2.4 (ii),

$$(3.14) \qquad P \leq \lambda_1(\Sigma^{-1})I_p.$$

Thus, (3.12) and Lemma 3.1 (i) imply the statement of the theorem. □

3.2. Upper bounds for $D[W^+]$. Put

$$H = \Sigma^{1/2}Y(Y'\Sigma Y)^{-1}(Y'\Sigma Y)^{-1}Y'\Sigma^{1/2}.$$

Now, similarly to (3.9),

$$E[\text{vec}W^+\text{vec}'W^+] = (2\pi)^{-np/2}\int e^{-\text{tr}\{YY'\}/2}\text{vec}H\text{vec}'H\,dY,$$

and performing the same transformations as when discussing $E[W^+]$, i.e. $Y \to (T, L)$ and thereafter $T \to V$ we end up with the following integral:

$$(3.15) \qquad \begin{aligned} E[\text{vec}W^+\text{vec}'W^+] = (2\pi)^{-np/2}2^{-n}\int e^{-\text{tr}\{V\}/2}|V|^{(p-n-1)/2} \\ \times (\Sigma^{1/2}L'(L\Sigma L')^{-1})^{\otimes 2}\text{vec}V^{-1}\text{vec}'V^{-1}((L\Sigma L')^{-1}L\Sigma^{1/2})^{\otimes 2}g(L)\,dLdV. \end{aligned}$$

Since by standardizing (3.15) appropriately we can assume that $V \sim W_n(I, p)$ and then, instead of (3.15), it follows from (1.6) by adding $E[\text{vec}W^+]E[\text{vec}'W^+]$ and using the definition of P given by (3.13), if $p-n-3 > 0$,

$$E[\text{vec}W^+\text{vec}'W^+] = c(n,p)^{-1}\int (c_1(I_{p^2} + K_{p,p})(P \otimes P) + c_2\text{vec}P\text{vec}'P)g(L)\,dL,$$

where

$$(3.16) \qquad c_1 = \frac{1}{(p-n)(p-n-1)(p-n-3)}, \quad c_2 = (p-n-2)c_1.$$

According to Definition 2.1 (iv) it is of interest to study, for an arbitrary α,

$$(3.17) \qquad (\alpha \otimes \alpha)'E[\text{vec}W^+\text{vec}'W^+](\alpha \otimes \alpha)$$

and

$$(3.18) \qquad E[(\alpha'W^+\alpha)^2] = c(n,p)^{-1}\int (2c_1 + c_2)(\alpha'P\alpha)^2 g(L)\,dL.$$

From Lemma 2.4 (i) and the inequality (3.14), upper and lower bounds of $E[(\alpha'W^+\alpha)^2]$ are obtained as follows:

THEOREM 3.4. *Let $W \sim W_p(\Sigma, n)$, $p > n + 3$, $\Sigma > 0$. For all $\alpha \in \mathbb{R}^p$,*

(i) $d(n,p)(\lambda_p(\Sigma^{-1}))^4(\alpha'\Sigma\alpha)^2 \leq E[(\alpha'W^+\alpha)^2] \leq d(n,p)(\lambda_1(\Sigma^{-1}))^4(\alpha'\Sigma\alpha)^2,$
(ii) $E[(\alpha'W^+\alpha)^2] \leq (2c_1 + c_2)(\lambda_1(\Sigma^{-1}))^2(\alpha'\alpha)^2,$

Shinpei Imori and Dietrich von Rosen   132

where $d(n,p) = (2c_1 + c_2)(2n + n^2)(2p + p^2)^{-1}$, c_1 and c_2 are defined in (3.16).

Proof. Combining (3.18) with Lemma 2.4 (i), we have

$$(2c_1 + c_2)(\lambda_p(\Sigma^{-1}))^4 \left\{ c(n,p)^{-1} \int (\alpha' \Sigma^{1/2} L' L \Sigma^{1/2} \alpha)^2 g(L)\, dL \right\}$$
$$\leq E[(\alpha' W^+ \alpha)^2]$$
$$\leq (2c_1 + c_2)(\lambda_1(\Sigma^{-1}))^4 \left\{ c(n,p)^{-1} \int (\alpha' \Sigma^{1/2} L' L \Sigma^{1/2} \alpha)^2 g(L)\, dL \right\}.$$

Hence, by applying Lemma 3.1 (iii), we can obtain (i). On the other hand, if we use the inequalities (3.14) instead of Lemma 2.4 (i), (3.18) implies that

$$E[(\alpha' W^+ \alpha)^2] \leq (2c_1 + c_2)(\lambda_1(\Sigma^{-1}))^2 (\alpha' \alpha)^2 \left\{ c(n,p)^{-1} \int g(L)\, dL \right\}.$$

Then, Lemma 3.1 (i) yields (ii).   □

Note that the exact value of $E[(\alpha' W^+ \alpha)^2]$ can be calculated when $\Sigma = I_p$ because according to Theorem 3.4 (i) the upper bound the equals lower bound, which is also verified from Proposition 1.1 (ii). Combining Theorem 3.4 with Theorem 3.2 or Theorem 3.3, upper bounds of $D[W^+]$ can be obtained.

THEOREM 3.5. *Let $W \sim W_p(\Sigma, n)$, $p > n + 3$, $\Sigma > 0$. According to Definition 2.1 (iv),*

$$D[W^+] \preceq \{ d(n,p)(\lambda_1(\Sigma^{-1}))^4 - a_1^2(\lambda_p(\Sigma^{-1}))^4 \}(\Sigma \otimes \Sigma),$$

$$D[W^+] \preceq (2c_1 + c_2)(\lambda_1(\Sigma^{-1}))^2 I_{p^2} - a_1^2(\lambda_p(\Sigma^{-1}))^4(\Sigma \otimes \Sigma),$$

where a_1 is given in Proposition 1.1 (i), $d(n,p)$ in Theorem 3.4, c_1 and c_2 are defined in (3.16).

It is worth noting that the inverse inequality of the first result in Theorem 3.5 is also established when $\Sigma = I_p$, which can be confirmed by Proposition 1.1. In this sense, the first upper bound is sharper than the second one if Σ is close to I_p. However, if $\lambda_1(\Sigma^{-1})$ is quite large (i.e., $\lambda_p(\Sigma)$ is very close to zero), then the second one may be better than the first one.

REFERENCES

[1] R.D. Cook and L. Forzani. On the mean and variance of the generalized inverse of a singular Wishart matrix. *Electron. J. Stat.*, 5:146–158, 2011.

[2] J.A. Díaz-García and R. Gutiérrez-Jáimez. Distribution of the generalised inverse of a random matrix and its applications. *J. Statist. Plann. Inference*, 136:183–192, 2006.

[3] T. Kollo and D. von Rosen. *Advanced Multivariate Statistics with Matrices.* Mathematics and Its Applications, Vol. 579, Springer, Dordrecht, 2005.

[4] T. Kollo, T. von Rosen, and D. von Rosen. Estimation in high-dimensional analysis and multivariate linear models. *Comm. Statist. Theory Methods*, 40:1241–1253, 2011.

[5] H. Neudecker and S. Liu. The density of the Moore-Penrose inverse of a random matrix. *Linear Algebra Appl.*, 237/238:123–126, Special issue honoring Calyampudi Radhakrishna Rao, 1996.

[6] R.F. Potthoff and S.N. Roy. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51:313–326, 1964.

[7] D. von Rosen. *Bilinear Regression Analysis. An Introduction.* Lecture Notes in Statistics, Vol. 220. Springer, New York, 2018.

[8] M.S. Srivastava and C.G. Khatri. *An Introduction to Multivariate Statistics.* North-Holland, New York, 1979.

 [9] T. Yamada, M. Hyodo, and T. Seo. The asymptotic approximation of EPMC for linear discriminant rules using a Moore-Penrose inverse matrix in high dimension. *Comm. Statist. Theory Methods*, 42:3329–3338, 2013.

[10] Y.T. Zhang. The exact distribution of the Moore-Penrose inverse of X with a density. *Multivariate Analysis VI (Pittsburgh, Pa., 1983)*, North-Holland, Amsterdam, 633–635, 1985.

[11] Z. Zhang. Pseudo-inverse multivariate/matrix-variate distributions. *J. Multivariate Anal.*, 98:1684–1692, 2007.