Estimation of normal covariance matrices parametrized by irreducible symmetric cones under Stein's loss

Yoshihiko Konno [1]

Faculty of Science, Japan Women's University, 2-8-2 Mejirodai Bunkyo-ku, Tokyo 112-8681, Japan

Abstract

In this paper the problem of estimating a covariance matrix parametrized by an irreducible symmetric cone in a decision-theoretic set-up is considered. By making use of some results developed in a theory of finite-dimensional Euclidean simple Jordan algebras, Bartlett's decomposition and an unbiased risk estimate formula for a general family of Wishart distributions on the irreducible symmetric cone are derived; these results lead to an extension of Stein's general technique for derivation of minimax estimators for a real normal covariance matrix. Specification of the results to the multivariate normal models with covariances which are parametrized by complex, quaternion, and Lorentz types, gives minimax estimators for each model.

Key words: Minimax, Stein estimator; Generalized Wishart distribution; Bartlett decomposition; Unbiased risk estimate; Jordan algebras.

1991 MSC: Primary 62H12; Secondary 62C20

1 Introduction

There has been considerable effort which has been expanded on constructing improved estimators for a covariance matrix of a multivariate normal distribu-

Email address: konno@fc.jwu.ac.jp (Yoshihiko Konno).
URL: http://mp-w3math.jwu.ac.jp/~konno/index.html (Yoshihiko Konno).

[1] This study was supported in part by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (C) (No.13680369 and No.17500185).

tion in order to obtain substantial reductions in risk since the pioneering paper of James and Stein [17]. They have considered the problem of estimating the covariance matrix under so-called Stein's loss function and obtained a minimax estimator with a constant risk, i.e., the best triangular-invariant estimator under the loss function. For the loss function, Stein [32] has given an unbiased risk estimate for a class of orthogonally invariant estimators, from which he obtained orthogonally invariant minimax estimators which are uniformly better than the best triangular-invariant estimator in [17]. At the same time, he also proposed so-called Stein's rough estimator and its modification which meets natural restrictions on the order of estimated eigenvalues. We note that there are three important ingredients among these works, i.e., an invariance argument on the triangular and orthogonal groups, Bartlett's decomposition, and the unbiased risk estimate for the class of orthogonally invariant estimators. The works mentioned above lead to the following natural question: Is it possible for any estimators to improve upon the maximum likelihood estimator for normal covariance matrices with complex or quaternion structure? The purpose of this paper is to obtain improved estimators for the covariance matrices under these models. To this end, we employ a unified method. The method involves an abstract theory of finite-dimensional Euclidean simple Jordan algebras, which has been extensively developed in [11]. Following the definition in [35], we consider zero-mean multivariate normal models with covariance matrices having some structures which are parametrized by one of the cones of the following four types: real, complex, quaternion, and Lorentz types. The merit of employing the theory of Euclidean Jordan algebras is to treat different four models in a unified manner. Then the maximum likelihood estimators under our normal models follow so-called general Wishart distributions on the symmetric cones; see, for examples, [2,5,7,8,23–25]. Using results developed in [11], we extend the important techniques for the estimation problem of the normal covariance matrix to construct improved estimators in our setting. The estimators obtained correspond to an orthogonally invariant minimax estimator which appeared in [9], and to Stein's rough estimator for the covariance matrix of real normal distribution. For the cases of the complex and quaternion covariance structures, improved minimax estimators are obtained by replacing orthogonal matrices with unitary matrices and by using a slight modification of the constants which appeared in the improved minimax estimators for the covariance matrix of the multivariate real normal distribution.

The organization of the paper is as follows: In Section 2 we describe principle aspects and some results of the indecomposable symmetric cones, almost equivalently those of finite-dimensional Euclidean simple Jordan algebras, which are taken from [11]. Furthermore we collect the results related to the general Wishart distributions on the symmetric cones, including Bartlett's decomposition of the general Wishart distributions. In Section 3, using the results stated in Section 2, we present main results of the paper. Using an approach due to [30], we obtain an unbiased risk estimate for a family of orthogonally invariant estimators in our setting. From these, we construct improved estimators which are counterparts in real normal setting. In Section 4, we demonstrate improved estimators in each of the models which include the normal models of the complex, quaternion, and Lorentz types.

In this paper, we extend the minimax estimation theory to the Wishart distributions on the symmetric cones. It is interesting to investigate whether this theory holds for the generalized Wishart distributions on homogeneous cones which were developed by Andersson and Wojnar [5]. Their models include the Wishart distributions which arise from group symmetry normal models and lattice conditional independence normal models as well as the Wishart distributions on the symmetric cones. There has not been literature on research of minimax estimation theory for these models except Konno [19] in which minimax estimators were obtained for the lattice conditional independence real normal models. It is interesting to develop a unified theory of the minimax estimation for more general models.

2 Preliminaries

In this section we first review some notions and known results for finite-dimensional Euclidean simple Jordan algebras. The material is taken from [11]. Next we introduce zero-mean multivariate normal distributions with covariance structures, i.e., real, complex, quaternion, and Lorentz types, and each covariance matrix in the models is parametrized by the corresponding symmetric cone. Furthermore we describe some results related to a general family of Wishart distributions which involve the estimation problem discussed in the sequel of this paper.

2.1 Some notions and known results for Euclidean Jordan algebras

A Jordan algebra $\mathcal{V}$ consists of a real vector space equipped with a bilinear product xy for any x and y in $\mathcal{V}$, satisfying the commutative law and the Jordan identity: $xy = yx$ and $(x^2y)x = x^2(yx)$. We only consider the theory of finite-dimensional Euclidean simple Jordan algebras; this theory is almost equivalent to that of indecomposable symmetric cones. See [11, Chapter II] for detailed definitions and properties of the Euclidean Jordan algebras. Among others, important facts and definitions we use in this paper are the following: There are five types of indecomposable symmetric cones: the three cones of Hermitian positive definite matrices with real, complex, and quaternion entries, the Lorentz cone, and the cone of 3×3 Hermitian positive definite matrices on the Cayley algebra. The Jordan algebra $\mathcal{V}$ is a power-associative algebra. We denote by e an identity element in $\mathcal{V}$. An element c in $\mathcal{V}$ is said to be idempotent if $c^2 = c$. Two idempotents c_1 and c_2 in $\mathcal{V}$ are said to be orthogonal if $c_1c_2 = 0$. An idempotent is said to be primitive if it is nonzero, and if it is not expressed as the sum of any two nonzero idempotents. A set of idempotents $\{c_1, c_2, \ldots, c_r\}$ is said to be a Jordan frame if it is a maximal system of orthogonal primitive idempotents, i.e., $c_ic_j = 0\,(i \neq j)$ and $\sum_{j=1}^{r} c_j = e$. It is known that any Jordan frame for a Jordan algebra $\mathcal{V}$ has the same number of elements, called the rank of $\mathcal{V}$. We denote by r the rank of the Jordan algebra $\mathcal{V}$. For any x in $\mathcal{V}$, there exist unique real numbers $x_1, x_2, \ldots, x_r$ and a Jordan frame $c_1, c_2, \ldots, c_r$ such that $x = \sum_{j=1}^{r} x_jc_j$. We define the trace and the determinant of the element x as $\mathrm{tr}(x) = \sum_{j=1}^{r} x_j$ and $\det(x) = \Pi_{j=1}^{r} x_j$. Without loss of generality, we can assume that an inner product on $\mathcal{V}$ is defined as $(x|\, y) = \mathrm{tr}(xy)$ for any x and y in $\mathcal{V}$. We use notation $\mathrm{Tr}(\,\cdot\,)$ and $\mathrm{Det}(\,\cdot\,)$ for the trace and the determinant of matrices and matrix representations of maps from a vector space to itself. An element x in $\mathcal{V}$ is said to be invertible if and only if $\det(x) \neq 0$. Here the invertibility means that there exists an element y that belongs to the polynomial algebra generated by x, such that $xy = e$. For a Jordan frame $c_1, c_2, \ldots, c_r$, set

$$\mathcal{V}_{ii} = \{x \in \mathcal{V}|\, c_ix = x\} \quad \text{and} \quad \mathcal{V}_{ij} = \{x \in \mathcal{V}|\, c_ix = (1/2)x,\, c_jx = (1/2)x\} \tag{1}$$

for $i = 1, 2, \ldots, r$ and $j = i+1, \ldots, r$. Then it is known that each of $\mathcal{V}_{ij}$'s $(i < j)$ has a common dimension g, called the Peirce invariant, and that $\mathcal{V}$ has the Peirce decomposition $\mathcal{V} = \oplus_{1\leq i\leq j\leq r}\mathcal{V}_{ij}$. Also we have

$$v = r + \frac{g}{2}r(r-1), \tag{2}$$

where v and r are the dimension and the rank of the Jordan algebra $\mathcal{V}$, respectively. We define the following maps: For elements x and y in $\mathcal{V}$, $L(x)y = xy$ and $P(x) = 2L(x)^2 - L(x^2)$. For an idempotent c and for an element z in $\{x \in \mathcal{V}|\, cx = (1/2)x\}$ we define the Frobenius transformation: $\tau_c(z) = \exp\{L(z) + 2L(z)L(c) - 2L(c)L(z)\}$. Here exponential map means that $\exp(A) = \sum_{j=0}^{\infty} A^j/j!$ for a map A. Let Ω be the corresponding symmetric cone associated to the Jordan algebra $\mathcal{V}$, i.e., $\Omega = \{x^2|\, x \in \mathcal{V}, \det(x) \neq 0\}$. Let $\mathrm{GL}(\mathcal{V})$ be the general linear group of $\mathcal{V}$. For an element h in $\mathrm{GL}(\mathcal{V})$ we denote by h^* the adjoint of h, i.e., $(hx|\, y) = (x|\, h^*y)$ for any x and y in $\mathcal{V}$. We denote by $\mathscr{G}(\Omega)$ the automorphism group of Ω defined by $\mathscr{G}(\Omega) = \{h \in \mathrm{GL}(\mathcal{V})|\, h\Omega = \Omega\}$, and we also denote by $\mathscr{G}$ the identity component of the automorphism group $\mathscr{G}(\Omega)$. We fix a Jordan frame as $c_1, c_2, \ldots, c_r$ in the remainder of this subsection. The triangular subgroup $\mathscr{T}$ of $\mathscr{G}$ is the set of an element t in $\mathscr{G}$ such that, for any x in $\mathcal{V}$,

$$(tx_{k\ell})_{ij} = \begin{cases} 0, & \text{if } (i, j) < (k, \ell), \\ \lambda_{ij}x_{ij}, & \text{if } (i, j) = (k, \ell), \end{cases}$$

where $x = \sum_{j=1}^{r} x_j c_j + \sum_{j<k} x_{jk}$ is the Peirce decomposition of x with respect to the Jordan frame $c_1, c_2, \ldots, c_r$. Here $x_1, x_2, \ldots, x_r$ are real numbers, $x_{k\ell} \in \mathcal{V}_{k\ell}\,(k < \ell)$, λ_{ij}'s are positive numbers, and $(i, j) < (k, \ell)$ means the lexicographic order.

The next lemma gives a collection of useful computational results on the Jordan algebra; these results are taken from [11].

Lemma 1 *Let $\mathcal{V}$ be a finite-dimensional Euclidean simple Jordan algebra of rank r, dimension v, and Peirce invariant g, and let Ω be the symmetric cone associated to the Jordan algebra $\mathcal{V}$. We have the following assertions.*

(i) *For an invertible element x in $\mathcal{V}$ we have $P(x)^{-1}x = x^{-1}$ and $P(x)^{-1} = P(x^{-1})$.*

(ii) *For an element x in $\mathcal{V}$ we have $\operatorname{Tr} L(x) = (v/r)\operatorname{tr}(x)$ and $\operatorname{Det} P(x) = (\det x)^{2v/r}$.*

(iii) *For an element u in Ω and an invertible element x in $\mathcal{V}$, we have*

$$\left.\frac{d}{dt}\right|_{t=0} \det(x+tu) = \det(x)(x^{-1}|\, u) \quad \textit{and} \quad \left.\frac{d}{dt}\right|_{t=0} (x+tu)^{-1} = -P(x)^{-1}u.$$

(iv) *For an element y in Ω there exists an element t in the triangular group $\mathscr{T}$ such that $y = t^*e$. Furthermore, we have $t^{-1}e = (t^*e)^{-1} = y^{-1}$.*

(v) *For an element x in Ω and t in $\mathscr{T}$ set $z = tx$. Then we have $\det(x) = \det(t^{-1}z) = \det(t^{-1}e)\det(z)$.*

(vi) *Let $\mathcal{V}_{ij}$ $(i = 1, 2, \ldots, r, j = i+1, \ldots, r)$ be defined by (1). For $i \neq j$ we have $\mathcal{V}_{ik}\mathcal{V}_{jk} \subset \mathcal{V}_{ij}$. If, in particular, $\{i, j\} \cap \{k, \ell\} = \emptyset$, then $\mathcal{V}_{ij}\mathcal{V}_{k\ell} = 0$.*

(vii) *Let c be an idempotent and let z be an element in $\mathcal{V}_{1/2} = \{y \in \mathcal{V} |\, cy = (1/2)y\}$. If $x = x_1 \oplus x_{12} \oplus x_0$ is the Peirce decomposition with respect to the idempotent c such that $x_1 \in \{y \in \mathcal{V} |\, cy = y\}$, $x_0 \in \{y \in \mathcal{V} |\, cy = 0\}$, and $x_{12} \in \mathcal{V}_{1/2}$, then we have*

$$\tau_c(z)(x) = x_1 \oplus (2L(z)x_1 + x_{12}) \oplus (2L(e-c)L(z)^2 x_1 + 2L(e-c)L(z)x_{12} + x_0).$$

(viii) *Let c_1 and c_2 be two orthogonal primitive nonzero idempotents. If a and b belong to $\mathcal{V}_{12} = \{y \in \mathcal{V} |\, c_1 y = (1/2)y$ and $c_2 y = (1/2)y\}$, then we have $ab = (1/2)(a|\, b)(c_1 + c_2)$.*

Proof. For the proof of (i)-(vii), see [11, Chapter II–VI]. The proof of (viii) can be obtained in a similar way as that of [11, Proposition IV.1.4]. □

It is known that the triangular group $\mathscr{T}$ has a parametrization as follows: Set $\mathcal{V}_+ = \{u \in \mathcal{V} |\, u = \sum_{i=1}^r u_i c_i + \sum_{i<j} u_{ij}, u_i > 0, u_{ij} \in \mathcal{V}_{ij}\}$ and for $u \in \mathcal{V}_+$ let

$$t(u) = P(b_1)\tau_{c_1}(u^{(1)})P(b_2)\tau_{c_2}(u^{(2)}) \times \cdots \times \tau_{c_{r-1}}(u^{(r-1)})P(b_r), \tag{3}$$

where $u^{(j)} = \sum_{k=j+1}^r u_{jk}$ and $b_j = c_1 + \cdots + c_{j-1} + u_j c_j + c_{j+1} + \cdots + c_r$. Then the map $u \mapsto t(u)$ from $\mathcal{V}_+$ onto $\mathscr{T}$ is a bijection.

The next lemma plays a vital role in obtaining a minimax risk for the estimation problem of this paper.

Lemma 2 *Assume that an element $t(u)$ in the triangular group $\mathscr{T}$ is given by* (3) *for an element $u \in \mathcal{V}_+$ such that $u = \sum_{j=1}^r u_j c_j + \sum_{j<k} u_{jk}$ is the Peirce decomposition of u with respect to the Jordan frame $c_1, c_2, \ldots, c_r$. We have three assertions.*

(i) *If $x = \sum_{j=1}^r x_j c_j + \sum_{j<k} x_{jk}$ is the Peirce decomposition of $x = t(u)\sum_{j=1}^r a_j c_j$ with $a_j > 0$ $(j = 1, 2, \ldots, r)$, then we have*

$$x_j = a_j u_j^2 + (1/2)\sum_{k=1}^{j-1} a_k \|u_{kj}\|^2 \qquad \textit{and} \qquad x_{jk} = a_j u_j u_{jk} + 2\sum_{\ell=1}^{j-1} a_\ell u_{\ell j} u_{\ell k}.$$

Furthermore, we have

$$\mathrm{tr}\,(x) = \sum_{j=1}^r a_j u_j^2 + (1/2)\sum_{j=1}^r \sum_{k=j+1}^r a_j \|u_{jk}\|^2,$$

where $\|\cdot\|$ is a Euclidean norm.

(ii) *If, in particular, $a_1 = a_2 = \cdots = a_r = 1$, then we have* $\det(x) = \Pi_{j=1}^r u_j^2$ *and* $\det(x)^{-v/r}\,dx = 2^r \Pi_{j=1}^r u_j^{-g(j-1)-1}\,du$, *where dx and du are the Lebesgue measures on $\mathcal{V}$ and $\mathcal{V}_+$, respectively.*

(iii) *If an element a_{ij} belongs to $\mathcal{V}_{ij}$ $(i < j)$, then we have*

$$\mathrm{tr}\,(t(u)a_{ij}) = u_j(u_{ij}|\, a_{ij}) + 2\sum_{\ell=j+1}^r (u_{i\ell}|\, u_{j\ell} a_{ij}).$$

Proof. The proof of the first part of (i) can be obtained by a minor modification of that of [11, Proposition VI.3.8]. The proof of second part of (i) follows from the first part of (i) and Lemma 1(ii). The proof of (ii) follows from [11, Theorem VI.3.9 and Proposition VI.3.10]. The proof of (iii) involves tedious calculation. First we observe that $P(b_\ell)a_{ij} = a_{ij}$ for $\ell \notin \{i, j\}$ and that $P(b_\ell)a_{ij} = u_\ell a_{ij}$ for $\ell \in \{i, j\}$. Furthermore we note that, from Lemma 1(vii), $\tau_{c_j}(u_{jk})y = y$ if y belongs to $\mathcal{V}_{mn}$ for $\{m, n\} \cap \{j, k\} = \emptyset$. From (3) and these observations we have that $t(u)a_{ij} = P(b_i)\tau_{c_i}(u^{(i)})P(b_j)\tau_{c_j}(u^{(j)})a_{ij}$, where

$\tau_{c_{r+1}}(u^{(r+1)}) = L(e)$, $c_{r+1} = 0$, and $u^{(r+1)} = 0$. Using Lemma 1(vi)-(viii) and noting the fact that $\mathrm{tr}\,(y) = 0$ if $y \in \mathcal{V}_{k\ell}$ $(k < \ell)$, we see that

$$\begin{aligned}
\mathrm{tr}\,(t(u)a_{ij}) &= \mathrm{tr}\,[P(b_i)\tau_{c_i}(u^{(i)})P(b_j)\{a_{ij} + 2\sum_{k=j+1}^{r} L(e-c_j)L(u_{jk})a_{ij}\}] \\
&= \mathrm{tr}\,[P(b_i)\tau_{c_i}(u^{(i)})\{u_j a_{ij} + 2\sum_{k=j+1}^{r} u_{jk}a_{ij}\}] \\
&= \mathrm{tr}\,[P(b_i)\{u_j a_{ij} + 2\sum_{k=j+1}^{r} u_{jk}a_{ij} \\
&\qquad +2\sum_{\ell=i+1}^{r} L(e-c_i)L(u_{i\ell})(u_j a_{ij} + 2\sum_{k=j+1}^{r} u_{jk}a_{ij})\}] \\
&= \mathrm{tr}\,[2\sum_{\ell=i+1}^{r} L(e-c_i)\{u_{i\ell}(u_j a_{ij}) + 4\sum_{k=j+1}^{r} u_{i\ell}(u_{jk}a_{ij})\}] \\
&= u_j(u_{ij}|\, a_{ij}) + 2\sum_{\ell=j+1}^{r} (u_{i\ell}|\, u_{j\ell}a_{ij}),
\end{aligned}$$

which completes the proof of (iii). □

To parametrize the covariance matrices in terms of the symmetric cones, we employ a symmetric representation as in [25,35]. Let $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$, where $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ stand for real, complex, and quaternion numbers, respectively. We also denote by $\mathbb{F}_H^{r\times r}$ and $\mathbb{F}_+^{r\times r}$ the spaces of $r \times r$ Hermitian matrices and Hermitian positive-definite matrices, respectively. Following [11], we say that $\mathbf{\Lambda}$ is a symmetric representation on $\mathbb{R}^p$ for appropriate dimension p if Λ is a linear map from a Jordan algebra $\mathcal{V}$ to the space of self-adjoint endomorphisms of $\mathbb{R}^p$ such that $\mathbf{\Lambda}(x^2) = \mathbf{\Lambda}(x)^2$ for any x in $\mathcal{V}$. Then we have basic properties on the symmetric representation $\mathbf{\Lambda}$ on $\mathbb{R}^p$ as follows.

Lemma 3 *Assume that $\mathcal{V}$ is a finite-dimensional Euclidean simple Jordan algebra with the rank r, dimension v, and Peirce invariant g. Let Ω be the symmetric cone associated to $\mathcal{V}$ and let $\mathbf{\Lambda}$ be a symmetric representation on $\mathbb{R}^p$. We have three assertions.*

(i) *For an element x in Ω, we have $\mathbf{\Lambda}(x^{-1}) = \mathbf{\Lambda}(x)^{-1}$.*

(ii) *For an element x in $\mathcal{V}$, we have $\mathrm{Det}\,\mathbf{\Lambda}(x) = \det(x)^{p/r}$.*

(iii) *Let x be an element in Ω. For any column vector ξ in $\mathbb{R}^p$ there exists an element y in the closure of the symmetric cone Ω such that $\xi'\mathbf{\Lambda}(x)\xi = (x|\, y)$, where ξ' stands for the transpose of a column vector ξ.*

Proof. The proof of (i) is obtained from the fact that $\mathbf{\Lambda}(e)$ is the identity transformation of $\mathbb{F}^p$, and that $\mathbf{\Lambda}(x)$ and $\mathbf{\Lambda}(x^{-1})$ are commutative. For the proof of (ii), see [11, Proposition IV.4.2] and [35, Lemma 2]. The assertion of (iii) is a well-known result which is obtained from the Reitz representation theorem since a map $x \mapsto \xi' \mathbf{\Lambda}(x) \xi$ is a linear map on $\mathcal{V}$. $\square$

2.2 Normal models parametrized by the symmetric cones and general Wishart distributions

We say that an $\mathbb{R}^p$-valued random column vector Z follows a multivariate normal distribution with zero-mean and a covariance matrix $\mathbf{\Lambda}(\sigma)$ for some σ in Ω if its density function with respect to the Lebesgue measure on $\mathbb{R}^p$ is

$$f_Z(\boldsymbol{z}) = (2\pi)^{-p/2} \operatorname{Det} \mathbf{\Lambda}(\sigma)^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{z}'\mathbf{\Lambda}(\sigma)^{-1}\boldsymbol{z}\right\}. \tag{4}$$

In the sequel of the paper, "$'$" stands for the transpose of a matrix. We denote by $\mathcal{N}_p(0,\, \mathbf{\Lambda}(\sigma))$ this distribution. As noted in [25], there are only four types of simple symmetric cones which parametrize the normal models defined by (4). The first three cones are the spaces of positive-definite Hermitian matrices with real, complex, and quaternion entries. The fourth one is the Lorentz cone, that is, $(\mathbb{R}\times\mathcal{W})_+ = \{(a_1,\, a_2) \in \mathbb{R}\times\mathcal{W} |\, a_1 > 0,\, a_1^2 - B(a_2|\, a_2) > 0\}$. Here we denote by $\mathbb{R}\times\mathcal{W}$ a vector space with multiplication defined by $(a_1,\, a_2)(b_1,\, b_2) = (a_1b_1 + B(a_2|\, b_2),\, a_1b_2 + b_1a_2)$ for $(a_1,\, a_2)$ and $(b_1,\, b_2) \in \mathbb{R}\times\mathcal{W}$, where $\mathcal{W}$ is a $(v-1)$-dimensional real vector space with a positive-definite bilinear form B on $\mathcal{W}$.

$\mathcal{V}$	p	Ω	$v = \dim_R(\mathcal{V})$	$r = \operatorname{rank}(\mathcal{V})$	g
$\mathbb{R}_H^{r\times r}$	r	$\mathbb{R}_+^{r\times r}$	$\frac{1}{2}r(r+1)$	r	1
$\mathbb{C}_H^{r\times r}$	$2r$	$\mathbb{C}_+^{r\times r}$	r^2	r	2
$\mathbb{H}_H^{r\times r}$	$4r$	$\mathbb{H}_+^{r\times r}$	$r(2r-1)$	r	4
$\mathbb{R}\times\mathcal{W}$	p	$(\mathbb{R}\times\mathcal{W})_+$	v	2	$v-2$

The investigation of normal models in a unified manner has been originated with [2] followed by [35]. For $\Omega = \mathbb{R}_+^{r\times r}$ and $p = r$, the density (4) becomes an r-variate real normal distribution, while, for $\Omega = \mathbb{C}_+^{r\times r}$ and $p = 2r$, it becomes a $2r$-variate real normal distribution with complex covariance structure which has been extensively studied by [1,14,18]. It is well-known that there is a one-to-one correspondence between an r-variate complex normal distribution and a $2r$-variate real normal distribution with complex covariance structure; this correspondence is established by an isomorphism between $\mathbb{C}^r$ and $\mathbb{R}^{2r}$. For $\Omega = \mathbb{H}_+^{r\times r}$ and $p = 4r$, the density (4) becomes a $4r$-variate real normal distribution with quaternion covariance structure. These three cases have been treated in a unified manner in [2]. For $\Omega = (\mathbb{R}\times\mathcal{W})_+$, the density (4) becomes the multivariate real normal distribution, the covariance matrix of which is parametrized by means of the Clifford algebras; this distribution has been appeared in [35].

Let $X_1, X_2, \ldots, X_n$ be a sequence of $\mathbb{R}^p$-valued random column vectors which are independently and identically distributed as a multivariate normal distribution $\mathcal{N}_p(0, \boldsymbol{\Lambda}(\sigma))$ for an element σ in Ω. Set $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ so $\boldsymbol{X}$ is a $p \times n$ random matrix. We denote by $\mathbb{R}^{p\times n}$ the space of $p \times n$ matrices with entries $\mathbb{R}$. Then $\boldsymbol{X}$ has its density function with respect to the Lebesgue measure

$$f_X(\boldsymbol{x}) = (2\pi)^{-np/2}\mathrm{Det}\,(\boldsymbol{\Lambda}(\sigma)\otimes \boldsymbol{I}_n)^{-1/2}\exp\left\{-\frac{1}{2}\mathrm{Tr}\,(\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\Lambda}(\sigma)^{-1})\right\}, \tag{5}$$

where $\otimes$ stands for the Kronecker product of matrices and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

Next we derive the maximum likelihood estimator for σ in (5) based on the random matrix $\boldsymbol{X}$.

From Lemma 3(iii) there exists an element w in the closure of the symmetric cone Ω such that

$$\mathrm{Tr}\,(\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Lambda}(y)) = (y|\, w) \tag{6}$$

for any y in $\mathcal{V}$. Note that w depends only on $\boldsymbol{X}$ so we write $w = Q(\boldsymbol{X})$. From Lemma 3(i) and (ii) we see that, given $\boldsymbol{X} = \boldsymbol{x}$, the likelihood function for σ is written as

$$\ell(\sigma|\,\boldsymbol{x}) = -\frac{np}{2r}\log\det(\sigma) - \frac{1}{2}(\sigma^{-1}|\, w) + (\text{a constant term}).$$

We use Lemma 1(i) and (iii) to see that the likelihood equation for σ is given by

$$\left.\frac{d}{dt}\right|_{t=0}\ell(\sigma + tu|\,\boldsymbol{x}) = -\frac{np}{2r}(\sigma^{-1}|\, u) + \frac{1}{2}(P(\sigma)^{-1}u|\, w) = 0, \tag{7}$$

for any u in $\mathcal{V}$. Because of positive-definiteness of the inner product $(\,\cdot\,|\,\cdot\,)$, the equation (7) implies $P(\sigma)^{-1}w = (np/r)\sigma^{-1}$. From Lemma 1(i) we have the solution for the equation (7) is given by $\sigma = (r/(np))w$. Therefore we write the maximum likelihood estimator for σ as $\widehat{\sigma}_{mle} = (r/(np))w$.

To derive the distribution of the maximum likelihood estimator for σ, we introduce general Wishart distributions on the symmetric cones. Assume that $N > v/r - 1$, which is the nonsingular part of the Gyndikin set; see, for example, [20]. We say that a random variable w in closure of the symmetric cone Ω follows a Wishart distribution on the symmetric cone Ω with parameter N and σ in Ω if w has a density with respect to the Lebesgue measure as follows:

$$f_\Omega(w|\, N,\, \sigma) = \frac{2^{-Nr}}{\Gamma_\Omega(N)} \det(\sigma)^{-N} \det(w)^{N-v/r} \exp\left\{-\frac{1}{2}(\sigma^{-1}|\, w)\right\}, \qquad (8)$$

where

$$\Gamma_\Omega(s) = (2\pi)^{(v-r)/2} \Pi_{j=1}^r \Gamma\left(s - \frac{g}{2}(j-1)\right).$$

Here $\Gamma(\,\cdot\,)$ is the usual gamma function. Note that, following the manner as in [11], the Lebesgue measure in (8) is chosen in order to correspond to the Euclidean structure $(x|\, y) = \mathrm{tr}\,(xy)$. Because of this, the factor $(2\pi)^{(v-r)/2}$ in $\Gamma_\Omega(s)$ replaces $\pi^{(v-r)/2}$ which appears in usual statistical literature. The notation for w having the Wishart distribution on the symmetric cone Ω with parameter N and σ in Ω is $\mathcal{JW}_\Omega(N,\, \sigma)$, briefly $\mathcal{L}(w) = \mathcal{JW}_\Omega(N,\, \sigma)$. Some fundamental properties of the general Wishart distributions on the indecomposable symmetric cones have been extensively investigated in [7,23,24]. More general extension of the classical Wishart distribution on the space of positive-definite real matrices, i.e., Wishart distributions on homogeneous cones, is provided in [5].

Proposition 4 *Let Ω be a symmetric cone of rank r, dimension v, and Peirce invariant g. Assume that a $p \times n$ random matrix $\boldsymbol{X}$ has the density function* (5)*, and let w be an element in the closure of the symmetric cone Ω such that $\mathrm{Tr}\,(\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Lambda}(y)) = (y|\, w)$ for any y in $\mathcal{V}$. Then we have $\mathcal{L}(w) = \mathcal{JW}_\Omega(np/(2r),\, \sigma)$.*

Proof. Let μ be the measure of $w = Q(\boldsymbol{X})$ on the closure of the symmetric cone Ω defined by

$$\int d\mu = (2\pi)^{-np/2}\, \mathrm{Det}\,(\boldsymbol{\Lambda}(\sigma) \otimes \boldsymbol{I}_n)^{-1/2} \int \exp\left\{-\frac{1}{2}\mathrm{Tr}\,(\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\Lambda}(\sigma)^{-1})\right\} d\boldsymbol{x}.$$

From Lemma 3(ii) and the definition of the symmetric representation $\boldsymbol{\Lambda}$ the Laplace transform of μ is

$$\int_\Omega e^{(y|\, w)}\, d\mu = (\mathrm{Det}\, \boldsymbol{\Lambda}(\sigma))^{-n/2} (\mathrm{Det}\,(\boldsymbol{\Lambda}(\sigma^{-1} - 2y))^{-n/2} = \det(e - 2\sigma y)^{-np/(2r)},$$

for an element y in $\mathcal{V}$, provided the integral above exists. On the other hand, the Laplace transform of $\mathcal{JW}_\Omega(N,\, \sigma)$ is

$$
\begin{aligned}
&\int_\Omega e^{(y|\, w)} f_\Omega(w|\, N,\, \sigma)\, dw \\
&\quad = \frac{2^{-Nr}}{\Gamma_\Omega(N)} \det(\sigma)^{-N} \int_\Omega \det(w)^{N-v/r} \exp\{-(\frac{\sigma^{-1}}{2} - y|\, w)\}\, dw \\
&\quad = 2^{-Nr} \det(\sigma)^{-N} \det(\frac{\sigma^{-1}}{2} - y)^{-N} \\
&\quad = \det(e - 2\sigma y)^{-N},
\end{aligned}
$$

for an element y in $\mathcal{V}$, provided the integral above exists. The second equality follows from [11, Proposition VII.1.2] and the last equality follows from Lemma 1(iv) and (v). Therefore these two equations imply the result of the theorem since the Laplace transform is injective. □

If we observe n copies of random vectors which are distributed as (4) and if we consider the problem of estimating σ based on these observations, then we have the induced problem of estimating σ based on w having the distribution $\mathcal{JW}_\Omega(N, \sigma)$ with $N = np/(2r)$. The next lemma is an elementary property of the general Wishart distributions.

Lemma 5 *Let w follow $\mathcal{JW}_\Omega(N, \sigma)$ distribution. Set $\sigma^{-1} = s^* e$ for an element s in the triangular group $\mathscr{T}$. If $z = sw$, then $\mathcal{L}(z) = \mathcal{JW}_\Omega(N, e)$.*

Proof. The proof can be obtained from Lemma 1(iv) and the fact that $\det(w)^{-v/r}\, dw$ is the $\mathscr{G}$-invariant measure. □

2.3 Bartlett's decomposition of the general Wishart distributions on symmetric cones

Throughout this subsection we assume that Ω is a symmetric cone of rank r, dimension v, and Peirce invariant g associated to a Jordan simple algebra $\mathcal{V}$ and that we fix a Jordan frame as $c_1, c_2, \ldots, c_r$.

Proposition 6 *Assume that $\mathcal{L}(w) = \mathcal{JW}_\Omega(N, e)$ and put $w = t(u)e$, where $t(u)$ is defined by (3) for an element $u \in \mathcal{V}_+$. If $u = \sum_{j=1}^r u_j c_j + \sum_{k>j} u_{jk}$ is the Peirce decomposition of u with respect to the Jordan frame $c_1, c_2, \ldots, c_r$, then we have that all elements of u are mutually independent with*

$$
u_j^2 \sim \chi^2_{2N-g(j-1)}, \qquad \textit{and} \qquad u_{jk} \sim \mathcal{N}_g(0,\, 2I_g)
$$

for $j = 1, 2, \ldots, r$ and $k = j+1, \ldots, r$.

Proof. The proof is obtained from an application of Lemma 2(i) and (ii). □

The next corollary is an immediate result from the Laplace transformation of the Wishart distribution and Proposition 6.

Corollary 7 *If* $\mathcal{L}(w) = \mathcal{J}W_\Omega(N,\, \sigma)$*, then* $\mathbb{E}[w] = 2N\sigma$ *and*

$$\mathbb{E}[\log\det(w)] - \log\det(\sigma) = \sum_{j=1}^{r} \mathbb{E}[\log u_j^2],$$

for $\mathcal{L}(u_j^2) = \chi^2_{2N-g(j-1)}$ $(j = 1,\, 2,\, \ldots\,,\, r)$.

3 Main results

Recall that Ω is the symmetric cone associated to a finite-dimensional simple Euclidean Jordan algebra $\mathcal{V}$ of rank r and Peirce invariant g. Assume that $\mathcal{L}(w) = \mathcal{J}W_\Omega(N,\, \sigma)$ for some σ in Ω and that $r \geq 2$ and $N > v/r$ in the sequel of the paper in order to guarantee to apply an integration-by-parts formula to the generalized Wishart distribution. Note that, from Proposition 4, the maximum likelihood estimator of σ is given by $\widehat{\sigma}_{mle} = w/(2N)$.

3.1 Loss function

We first derive a loss function based on Kullback-Leibler information. To this end, assume that we observe a random variable w from $\mathcal{J}W_\Omega(N,\, \sigma)$ and that $\widehat{\sigma}$ is an estimator for σ based on w. Given w, consider Kullback-Leibler distance from an estimated density $f_\Omega(\cdot\,|\,N,\, \widehat{\sigma})$ to the true density $f_\Omega(\cdot\,|\,N,\, \sigma)$, i.e.,

$$\mathrm{KL}(f_\Omega(\cdot\,|\,N,\, \widehat{\sigma}),\, f_\Omega(\cdot\,|\,N,\, \sigma)) = \int_\Omega f_\Omega(\tilde{w}|\,N,\, \widehat{\sigma}) \log\left(\frac{f_\Omega(\tilde{w}|\,N,\, \widehat{\sigma})}{f_\Omega(\tilde{w}|\,N,\, \sigma)}\right) d\tilde{w}.$$

Then we have

$$\begin{aligned}
&\mathrm{KL}(f_\Omega(\cdot\,|\,N,\, \widehat{\sigma}),\, f_\Omega(\cdot\,|\,N,\, \sigma)) \\
&\quad = \mathbb{E}\left[\frac{1}{2}(\sigma^{-1}|\,\tilde{w}) - \frac{1}{2}(\widehat{\sigma}^{-1}|\,\tilde{w}) - N\log\det(\widehat{\sigma}) + N\log\det(\sigma)\right] \\
&\quad = N\Big[(\sigma^{-1}|\,\widehat{\sigma}) - \log\det(\widehat{\sigma}) + \log\det(\sigma) - r\Big],
\end{aligned}$$

where the expectation above is taken with respect to $\mathcal{J}W_\Omega(N,\, \widehat{\sigma})$. The last equality follows from the fact that $\mathbb{E}[\tilde{w}] = 2N\widehat{\sigma}$ since $\tilde{w}$ follows the distribution $\mathcal{J}W_\Omega(N,\, \widehat{\sigma})$ and that $(\widehat{\sigma}^{-1}|\,\widehat{\sigma}) = (e|\,e) = \mathrm{tr}(e) = r$. Therefore, this suggests employing a loss function

$$\mathscr{L}(\widehat{\sigma},\, \sigma) = (\sigma^{-1} |\, \widehat{\sigma}) - \log\det(\widehat{\sigma}) + \log\det(\sigma) - r, \tag{9}$$

where $\widehat{\sigma}$ is an estimator for σ. Note that $\mathscr{L}(\cdot\,,\,\cdot)$ is a strictly convex of its first argument and that it is nonnegative and minimized at $\widehat{\sigma} = \sigma$ as usual. The loss function is a counterpart of the usual Stein loss function for the problem of estimating a normal covariance matrix. The risk function is defined as

$$\mathscr{R}(\hat{\sigma},\, \sigma) = \mathbb{E}[(\sigma^{-1} |\, \widehat{\sigma}) - \log\det(\hat{\sigma}) + \log\det(\sigma) - r],$$

where the expectation above is taken with respect to $\mathcal{J}W_{\Omega}(N,\, \sigma)$.

3.2 Minimax risk

From Corollary 7 we can see that the maximum likelihood estimator $\widehat{\sigma}_{mle} = w/(2N)$ is unbiased for σ. If we consider a class of estimator of the form $\widehat{\sigma}_{\alpha} = \alpha w$ for a positive constant α, then the unbiased estimator $\widehat{\sigma}_{mle}$ is the best among the class.

Proposition 8 *Let $\mathcal{L}(w) = \mathcal{J}W_{\Omega}(N,\, \sigma)$ and consider the estimators of the form αw for a positive constant α. Then the best constant multiple of αw is given by $\alpha = 1/(2N)$ with its risk*

$$\mathscr{R}((2N)^{-1}w,\, \sigma) = r\log(2N) - \sum_{j=1}^{r} \mathbb{E}[\log u_j^2],$$

for $\mathcal{L}(u_j^2) = \chi^2_{2N-g(j-1)}$ $(j = 1,\, 2,\, \ldots\,,\, r)$.

Proof. From Corollaries 7 and the fact that $(\sigma^{-1} |\, \sigma) = \mathrm{tr}(e) = r$ we have

$$\mathscr{R}(\widehat{\sigma}_{\alpha},\, \sigma) = 2\alpha N r - r\log\alpha - \sum_{j=1}^{r} \mathbb{E}[\log u_j^2] - r.$$

Differentiating the right hand side of above equation with respect to α, we can see that $\mathscr{R}(\widehat{\sigma}_{\alpha},\, \sigma)$ is minimized when $\alpha = 1/(2N)$. □

We consider a class of estimators satisfying

$$\widehat{\sigma}(sw) = s\widehat{\sigma}(w) \tag{10}$$

for any element s in the triangular group $\mathscr{T}$. Standard arguments such as those in [27, Section 4.3] and [10, Section 6.3], show that (10) holds if and only if

$$\hat{\sigma}_\delta(w) = t(u)\Bigl(\sum_{i=1}^{r} \delta_i c_i + \sum_{i<j} \delta_{ij}\Bigr) \qquad \text{and} \qquad w = t(u)e, \tag{11}$$

where $t(u)$ is uniquely determined as an element in $\mathscr{T}$ and defined by (3) with the element u in $\mathcal{V}_+$, and where δ_j's ($j = 1, 2, \ldots, r$) are positive constants and δ_{ij}'s ($i < j$) are constant elements in $\mathcal{V}_{ij}$ so that $\sum_{i=1}^r \delta_i c_i + \sum_{i<j} \delta_{ij} \in \Omega$. In fact, we note that $\widehat{\sigma}(e) = (t(u))^{-1}\widehat{\sigma}(w)$ if we set $s^{-1} = t(u)$ with $w = t(u)e$ and that $\widehat{\sigma}(e)$ in Ω is a constant.

Proposition 9 *Consider the estimators $\widehat{\sigma}_\delta(w)$ given by* (11) *and let u_j^2's ($j = 1, 2, \ldots, r$) be random variables, each of which is distributed as $\mathcal{L}(u_j^2) = \mathcal{L}(\chi^2_{2N-g(j-1)})$. Then we have*

$$\mathscr{R}(\widehat{\sigma}_\delta, \sigma) \geq \sum_{j=1}^{r} \Bigl\{ \log\Bigl(2N + g(r - 2j + 1)\Bigr) - \mathbb{E}[\log u_j^2] \Bigr\},$$

where the equality holds if and only if

$$\delta_i = \Bigl(2N + g(r - 2j + 1)\Bigr)^{-1} \qquad \textit{and} \qquad \delta_{ij} = 0 \tag{12}$$

for $i = 1, 2, \ldots, r$ and $j = i + 1, \ldots, r$. Furthermore, an estimator $\widehat{\sigma}_m = t(u)\sum_{j=1}^r \delta_j c_j$ with δ_j's, being given by (12), *is minimax.*

Proof. Set $\delta = \sum_{j=1}^r \delta_j c_j + \sum_{i<j} \delta_{ij}$. Note that $\det(\delta) \neq 0$. First we show that we may assume that $\sigma = e$ without loss of generality. To this end, write $\sigma^{-1} = s^* e$ for an element s in the triangular subgroup $\mathscr{T}$ and set $z = sw$. From Lemma 5 we have $\mathcal{L}(z) = \mathcal{J}W_\Omega(N, e)$. From (10) we have

$$\begin{aligned}\mathbb{E}[(\sigma^{-1}|\,\widehat{\sigma}_\delta(w))] &= \mathbb{E}[(s^*e|\,\widehat{\sigma}_\delta(w))] = \mathbb{E}[(e|\,s\widehat{\sigma}_\delta(w))] = \mathbb{E}[(e|\,\widehat{\sigma}_\delta(sw))] \\ &= \mathbb{E}[(e|\,\widehat{\sigma}_\delta(z))].\end{aligned}$$

Furthermore we can see that, from Lemma 1(iv) and (v),

$$\mathbb{E}[\log\det\widehat{\sigma}_\delta(w)] = \mathbb{E}[\log\det\widehat{\sigma}_\delta(z)] + \log\det\sigma.$$

From these two equations we can assume that $\sigma = e$ without loss of generality. To evaluate the risk of the estimator $\widehat{\sigma}_\delta(w)$ in the case that $\mathcal{L}(w) = \mathcal{J}W_\Omega(N, e)$, we use Lemma 2(i), (iii), and Proposition 6 to get that

$$
\begin{aligned}
\mathbb{E}[(e|\,\widehat{\sigma}_\delta(w))] &= \sum_{j=1}^{r} \mathbb{E}[\mathrm{tr}\{t(u)(\delta_j c_j)\}] + \sum_{i<j} \mathbb{E}[\mathrm{tr}\{t(u)\delta_{ij}\}] \\
&= \sum_{j=1}^{r} \Big\{ \delta_j \mathbb{E}[u_j^2] + \frac{1}{2} \sum_{k=1}^{j-1} \delta_k \mathbb{E}\Big[\|u_{kj}\|^2\Big] \Big\} \\
&= \sum_{j=1}^{r} \Big\{ \Big(2N - g(j-1)\Big)\delta_j + g \sum_{k=1}^{j-1} \delta_k \Big\} \\
&= \sum_{j=1}^{r} \Big(2N + g(r - 2j + 1)\Big)\delta_j.
\end{aligned} \tag{13}
$$

Furthermore we use Lemma 1(v) and Lemma 2(ii) to see that

$$
\begin{aligned}
\mathbb{E}[\log\det(\widehat{\sigma}_\delta(w))] &= \mathbb{E}[\log\det(t(u)e)] + \log\det\delta \\
&= \mathbb{E}\Big[\sum_{j=1}^{r} \log u_j^2\Big] + \log\det\delta.
\end{aligned} \tag{14}
$$

Hence combination of the equations (13) and (14) shows that the risk for the estimators (11) is given by

$$
h(\delta) = \mathscr{R}(\widehat{\sigma}_\alpha,\, e) = \sum_{j=1}^{r} \eta_j (c_j|\,\delta) - \log\det\delta - \sum_{j=1}^{r} \mathbb{E}[\log u_j^2] - r
$$

for $\eta_j = 2N + g(r - 2j + 1)$ $(j = 1, 2, \ldots, r)$. We differentiate $h(\delta)$ toward an element u in $\mathcal{V}$, equate it to zero, and use Lemma 1(iii) to get

$$
\frac{d}{dt}\bigg|_{t=0} h(\delta + tu) = \sum_{j=1}^{r} \eta_j (c_j|\,u) - (\delta^{-1}|\,u) = 0,
$$

for any u in $\mathcal{V}$. Since the positive-definiteness of the inner product, we conclude that $\delta^{-1} = \sum_{j=1}^{r} \eta_j c_j$, which shows that $\delta_j = \eta_j^{-1}$ and $\delta_{ij} = 0$ $(i = 1, 2, \ldots, r$ and $j = i+1, , \ldots, r)$. Furthermore, we can see that the lower bound for the risk of the estimators (11) is given by $h(\sum_{j=1}^{r} \eta_j^{-1} c_j) = \sum_{j=1}^{r} \{\log \eta_j - \mathbb{E}[\log u_j^2]\}$. This gives the proof of the first assertion of the theorem. Since the subgroup $\mathscr{T}$ is solvable, we apply the Hunt-Stein theorem (see [6] for the details) to complete the proof of the theorem. □

From Proposition 9, the estimator $t(u) \sum_{j=1}^{r} \eta_j^{-1} c_j$ is the best among the class of estimators (11). In the sequel, we set $\delta_j = \eta_j^{-1}$ $(j = 1, 2, \ldots, r)$ and denote by $\widehat{\sigma}_\delta$ the estimator $t(u) \sum_{j=1}^{r} \delta_j c_j$, such that $w = t(u)e$ and $t(u)$ is given by (3). We also observe that, by Jensen's inequality and the concavity of the logarithmic function, $\mathscr{R}((2N)^{-1}w,\, \sigma) \geq \mathscr{R}(\widehat{\sigma}_\delta,\, \sigma)$, in which the inequality holds strictly uniformly in any σ in Ω.

Recall that $\mathscr{G}$ is the connected component of the identity in the automorphism group of Ω. Let $\mathscr{O}(\mathcal{V})$ be the orthogonal group of $\mathcal{V}$. Also let $\mathscr{K} = \mathscr{G} \cap \mathscr{O}(\mathcal{V})$. We fix a Jordan frame $c_1, c_2, \ldots, c_r$. From [11, Corollary IV.2.7], we can see that any $w \in \Omega$ is written as $w = ka, k \in \mathscr{K}$, and $a = \sum_{j=1}^r a_j c_j$ for $a \in \mathcal{R}_+ = \{a = \sum_{j=1}^r a_j c_j | \, a_1 > a_2 > \cdots > a_r\}$. Set $\boldsymbol{a} = (a_1, a_2, \ldots, a_r)$.

Consider a family of estimators of the form

$$\widehat{\sigma}_\varphi = k \sum_{j=1}^{r} \varphi_j(\boldsymbol{a}) c_j, \tag{15}$$

in which $\varphi_j(\boldsymbol{a})$'s $(j = 1, 2, \ldots, r)$ are real-valued functions of $\boldsymbol{a}$. We write $\varphi_j(\boldsymbol{a})$ as φ_j. To evaluate the risk of estimators being given by (15), an identity in the next lemma plays a fundamental role.

Lemma 10 *Assume that $\mathcal{L}(w) = \mathcal{J}W_\Omega(N, \sigma)$ for some σ in Ω and consider the estimators $\widehat{\sigma}_\varphi(w)$ given by* (15)*. Under certain regularity conditions stated in* [30] *and which hold here, we have*

$$\mathbb{E}[(\sigma^{-1} | \, \widehat{\sigma}_\varphi)] = \mathbb{E}\left[\sum_{j=1}^{r}\left\{2\frac{\partial \varphi}{\partial a_j} + \left(2N - \frac{2v}{r}\right)\frac{\varphi_j}{a_j} + 2g\sum_{i=1}^{j-1}\frac{\varphi_i - \varphi_j}{a_i - a_j}\right\}\right].$$

Proof. Let

$$F(\boldsymbol{a}) = K \det(\sigma)^{-N}\left(\Pi_{j=1}^r a_j\right)^{N-v/r} \Pi_{j>i}(a_i - a_j)^g \text{ with } K = 2^{-Nr}/\Gamma_\Omega(N),$$
$$G(\boldsymbol{a}) = \int_{\mathscr{K}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{r} a_j (k^* \sigma^{-1} | \, c_j)\right\} dk,$$

where dk is the normalized Haar measure of $\mathscr{K}$. From [11, page 104], the joint density of $\boldsymbol{a} = (a_1, a_2, \ldots, a_r)$ with respect to the Lebesgue measure da_j $(j = 1, 2, \ldots, r)$ is given by $F(\boldsymbol{a})G(\boldsymbol{a})$. To derive an integration-by-parts formula, we set

$$\mathcal{R}_j = \{\boldsymbol{a}^{(j-)} = (a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_r) | \; a_1 > \cdots > a_{j-1} > a_{j+1} > \cdots > a_r > 0\}$$

with $a_0 = \infty$ and $a_{r+1} = 0$. From a direct calculation which involves integration-by-parts, we have

$$
\begin{aligned}
\mathbb{E}[(\sigma^{-1}|\,\widehat{\sigma}_\varphi)] &= \sum_{j=1}^{r} \mathbb{E}[\varphi_j\,(k^*\sigma^{-1}|\,c_j)] \\
&= -2\sum_{j=1}^{r}\int_{\mathcal{R}_j}\left(\int_{a_{j+1}}^{a_{j-1}} \varphi_j\, F(\boldsymbol{a})\frac{\partial}{\partial a_j}G(\boldsymbol{a})\,da_j\right)d\boldsymbol{a}^{(j-)} \\
&= 2\sum_{j=1}^{r}\int_{\mathcal{R}_j}\int_{a_{j+1}}^{a_{j-1}}\frac{\partial}{\partial a_j}(\varphi_j F(\boldsymbol{a}))G(\boldsymbol{a})\,da_j\,d\boldsymbol{a}^{(j-)} \\
&= 2\mathbb{E}\left[\sum_{j=1}^{r}\frac{\partial}{\partial a_j}(\varphi_j F(\boldsymbol{a}))\frac{1}{F(\boldsymbol{a})}\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{r}\left\{2\frac{\partial\varphi_j}{\partial a_j} + \left(2N - \frac{2r}{r}\right)\frac{\varphi_j}{a_j} + 2g\sum_{i\neq j}\frac{\varphi_j}{a_j - a_i}\right\}\right],
\end{aligned}
$$

which completes the proof of the lemma. □

Let $\mathscr{R}^{\#}(\widehat{\sigma}_\varphi,\,\sigma) = \mathbb{E}[(\sigma^{-1}|\,\widehat{\sigma}_\varphi) - \log\det(\widehat{\sigma}_\varphi)]$. It is easily seen that comparison between two estimators of the form (15) in terms of the risk $\mathscr{R}$ is equivalent to that in terms of $\mathscr{R}^{\#}$. The next theorem is a generalization of [9, Lemma 2.1] to the setting of general Wishart distributions on the symmetric cones.

Theorem 11 *Consider the estimators given by* (15)*. Then an unbiased risk estimate for* $\mathscr{R}^{\#}(\widehat{\sigma}_\varphi,\,\sigma)$ *is given as*

$$
\widehat{\mathscr{R}}^{\#}(\widehat{\sigma}_\varphi) = \sum_{j=1}^{r}\left\{2\frac{\partial\varphi}{\partial a_j} + \left(2N - \frac{2v}{r}\right)\frac{\varphi_j}{a_j} + 2g\sum_{i=1}^{j-1}\frac{\varphi_i - \varphi_j}{a_i - a_j} - \log\varphi_j\right\}, \qquad (16)
$$

i.e., we have $\mathscr{R}^{\#}(\widehat{\sigma}_\varphi,\,\sigma) = \mathbb{E}[\widehat{\mathscr{R}}^{\#}(\widehat{\sigma}_\varphi)]$.

Proof. The proof of the theorem follows from Lemma 10 and the fact that $\det(\,\cdot\,)$ is invariant under the transformation of $\mathscr{K}$. □

Next we consider a special form of the estimators (15), i.e., estimators of the form

$$
\widehat{\sigma}_m = k\sum_{j=1}^{r}\varphi_j(\boldsymbol{a})c_j \qquad \text{and} \qquad \varphi_j(\boldsymbol{a}) = \delta_j a_j. \qquad (17)
$$

Here δ_j's ($j = 1,\,2,\,\ldots\,,\,r$) are given by (12), and $w = ka$ for $a = \sum_{j=1}^{r} a_j c_j$ in $\mathcal{R}_+$ and k in $\mathscr{K}$. The next theorem, which extends the result given in [9, Theorem 3.1] to the case of the symmetric cones, easily follows from Theorem 11.

Proposition 12 *Consider the estimator $\widehat{\sigma}_m$ given by* (17). *Then we have*

$$\mathscr{R}(\widehat{\sigma}_m,\, \sigma) \leq -\sum_{j=1}^{r} \log \delta_j - \sum_{j=1}^{r} \mathbb{E}[\log u_j^2],$$

where u_j^2's follow $\chi^2_{2N-g(j-1)}$ $(j = 1,\, 2,\, \dots\,,\, r)$. Hence the estimator $\widehat{\sigma}_m$, being given by (17)*, is minimax.*

Proof. We note the fact that $\det(\,\cdot\,)$ is invariant under automorphism (see [11, Proposition II.4.2]) and use Corollary 7 to get that $\mathbb{E}[\log\det\widehat{\sigma}_m] = \sum_{j=1}^{r}\Big\{\log\delta_j + \mathbb{E}[\log u_j^2]\Big\} + \log\det\sigma$. From this equation and an application of Theorem 11 with $\varphi_j = \delta_j a_j$ $(j = 1,\, 2,\, \dots\,,\, r)$ we have

$$\begin{aligned}
\mathscr{R}(\widehat{\sigma}_{\varphi^\delta},\, \sigma) &= \mathbb{E}\Bigg[2g\sum_{j=1}^{r}\sum_{i>j}\frac{\delta_j a_j - \delta_i a_i}{a_j - a_i} + 2\sum_{j=1}^{r}\delta_j + \left(2N - \frac{2v}{r}\right)\sum_{j=1}^{r}\delta_j \\
&\qquad - \sum_{j=1}^{r}\log u_j^2 - \sum_{j=1}^{r}\log\delta_j - r\Bigg] \\
&\leq \sum_{j=1}^{r}\left(2N - \frac{2(v-r)}{r} + 2g(r-j)\right)\delta_j - \sum_{j=1}^{r}\mathbb{E}[\log u_j^2] - \sum_{j=1}^{r}\log\delta_j - r \\
&= -\sum_{j=1}^{r}\mathbb{E}[\log u_j^2] - \sum_{j=1}^{r}\log\delta_j.
\end{aligned}$$

The second inequality follows from the same argument as the proof of [9, Theorem 3.1] and the last equality follows from the fact that $2(v-r)/r = g(r-1)$ by (2). □

Besides the estimator (17), from some applications of Theorem 11, we can obtain other minimax estimators. For example, following the approach in [9], we can improve upon the estimator (17). Furthermore, following the approach in [28], we can modify the estimator (17) to obtain an ordered preserving estimator, i.e., estimators satisfying a natural constraint $\varphi_1 \geq \varphi_2 \geq \cdots \geq \varphi_r$ of the eigenvalues of the orthogonally invariant estimators.

Here we derive the so-called Stein rough estimator following the approach in [30]. For the estimators of the form (15), note that, omitting constant terms which do not depend on the form of the estimators (15), the loss function

$$\mathscr{L}^{\#}(\widehat{\sigma}_\varphi,\, \sigma) = \sum_{j=1}^{r}\varphi_j(\boldsymbol{a})(k^*\sigma^{-1}|\, c_j) - \sum_{j=1}^{r}\log\varphi_j(\boldsymbol{a})$$

is minimized at $\varphi_j(\boldsymbol{a}) = 1/(k^*\sigma^{-1}|\, c_j)$ $(j = 1, 2, \dots\,, r)$. However, it depends on the unknown parameter σ. So we replace $(k^*\sigma^{-1}|\, c_j)$ with its unbiased estimator $\widehat{\alpha}_j$, i.e., $\mathbb{E}[\widehat{\alpha}_j] = \mathbb{E}[(k^*\sigma^{-1}|\, c_j)]$, which gives an estimator

$k\sum_{j=1}^{r} c_j/\widehat{\alpha}_j$. In fact, following the approach in [30], we have $\mathbb{E}[(k^*\sigma^{-1}|\, c_j)] = \mathbb{E}[2(\partial/\partial a_j)\log F(\boldsymbol{a})]$, where $F(\boldsymbol{a})$ is given in the proof of Lemma 10. This results in, for $j = 1, 2, \ldots, r$,

$$\widehat{\alpha}_j = 2\frac{\partial}{\partial a_j}\log F(\boldsymbol{a}) = \left(2N - \frac{2v}{r}\right)\frac{1}{a_j} + 2g\sum_{\ell\neq j}\frac{1}{a_j - a_\ell}.$$

Hence we can obtain the estimator given by

$$\widehat{\sigma}_{st} = k\sum_{j=1}^{r}\frac{a_j}{2N - 2v/r + 2g\sum_{\ell\neq j} a_j/(a_j - a_\ell)}c_j, \tag{18}$$

which generalizes Stein's rough estimator. The estimators $\widehat{\alpha}_j$ cannot be used directly because they frequently do not satisfy order restrictions $\widehat{\alpha}_1 \geq \widehat{\alpha}_1 \geq \cdots \geq \widehat{\alpha}_r$, and some of them might be negative. Therefore, Stein [32] recommended an isotonizing adjustment; see [21] for further details. Similarly the estimator (16) does not satisfy natural constraints, i.e., $\varphi_1(\boldsymbol{a}) \geq \varphi_2(\boldsymbol{a}) \geq \cdots \geq \varphi_r(\boldsymbol{a})$. Note that, if an orthogonally invariant estimator for a covariance matrix does not satisfy the natural constraints, then it can be dominated in terms of risk by its modifications which preserve the order. Therefore some correction to meet the natural constraints is necessary in a manner such that methods are described in [16,21,28,31].

4 Examples of Improved Estimators under specified normal models

In this section we develop improved estimation method in examples of multivariate normal distributions which include the complex, quaternion, and Lorentz types. Tolver Jensen [35] proved that a normal model in the case that the covariance hypotheses is linear in both the covariance and the inverse covariance can be represented as a product of normal models; each factor in the product is one of the following four types: independent repetitions of a real normal model, independent repetitions of a complex normal model, independent repetitions of a quaternion normal model, and independent repetitions of a Lorentz normal model.

The complex normal and the complex Wishart distributions are well-known statistical models. See, for example, [1,4,14,15,18,29]. The multivariate complex normal models are often used for the description of statistical properties in a common problem in signal processing. The knowledge of the complex covariance matrix is closely related to typical parameters of interest in this type of the problem. See [22] for this. Recently Svensson and Lundberg [33] considered the problem of estimating the complex covariance matrix in a decision-theoretic manner and obtained the same results as those presented in Section 4.1 via a different method. A zero-mean quaternion normal model has appeared in [2] as an example of invariant normal models. However, there has been no literature on investigating estimation problem of the covariance matrix of the quaternion normal model in the decision-theoretic point of view as far as we know. Minimax estimation theory obtained in Section 4.2 is new. The Lorentz normal models discussed in Section 4.3 are peculiar and have different features from those discussed in Sections 4.1 and 4.2. We present the results of the minimax estimation theory on the Lorentz normal models because of their interesting and unusual features.

4.1 Complex Normal distributions

Let $\mathbb{C}_H^{r\times r}$ be the vector space of Hermitian $r\times r$ matrices with complex entries and define

$$\boldsymbol{C}_1 \circ \boldsymbol{C}_2 = \frac{1}{2}(\boldsymbol{C}_1\boldsymbol{C}_2 + \boldsymbol{C}_2\boldsymbol{C}_1), \tag{19}$$

for $\boldsymbol{C}_1$ and $\boldsymbol{C}_2 \in \mathbb{C}_H^{r\times r}$ where $\boldsymbol{C}_1\boldsymbol{C}_2$ is the usual matrix multiplication. Then it is known that $\mathbb{C}_H^{r\times r}$ with the multiplication $\circ$ is a Euclidean simple Jordan algebra with the rank r and that the trace, determinant and inverse in the sense of Jordan algebras are the usual ones. For $\mathbb{C}_H^{r\times r}$, we fix a Jordan frame $c_1, c_2, \dots, c_r$, each of which is an $r \times r$ diagonal matrix with j-th ($j = 1, 2, \dots, r$) element being one and the other elements being zero. An element $\boldsymbol{B}$ in the general linear group, the space of $r\times r$ nonsingular matrices with complex entries, acts on an element $\boldsymbol{C}$ in $\mathbb{C}_H^{r\times r}$ as $\boldsymbol{C} \mapsto \boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^*$. For any $\boldsymbol{C} \in \mathbb{C}^{q\times r}$ and $\boldsymbol{c} \in \mathbb{C}^r$, write $\boldsymbol{C} = \boldsymbol{A}+\boldsymbol{i}\boldsymbol{B}$ and $\boldsymbol{c} = \boldsymbol{a}+\boldsymbol{i}\boldsymbol{b}$, where $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{q\times r}$, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^r$, and $\boldsymbol{i}$ is the imaginary unit. Then we define the partitioned vector $[\boldsymbol{c}] \in \mathbb{R}^{2r}$ and the matrix $\{\boldsymbol{C}\} \in \mathbb{R}^{2q\times 2r}$ as

$$[\boldsymbol{c}] = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}, \qquad \{\boldsymbol{C}\} = \begin{pmatrix} \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{pmatrix}.$$

We also define the representation of $\boldsymbol{\Lambda}$ of $\mathbb{C}_H^{r\times r}$ in $\mathbb{R}^{2r}$ by $\boldsymbol{\Lambda}(\boldsymbol{C})[\boldsymbol{x}] = [\boldsymbol{C}\boldsymbol{x}]$ for any $\boldsymbol{x} \in \mathbb{C}^r$ and $\boldsymbol{C} \in \mathbb{C}_H^{r\times r}$. From [23, page 380] and Lemma 3(ii), we have $\mathrm{Tr}\{\boldsymbol{C}\} = 2\mathrm{tr}\,\boldsymbol{C}$ for any $\boldsymbol{C} \in \mathbb{C}_H^{r\times r}$. From this fact, (6) and (19), the corresponding $\mathbb{C}_H^{r\times r}$-valued quadratic form is given by $Q(\boldsymbol{X}) = (1/2)\sum_{i=1}^n X_i X_i^*$ for $\boldsymbol{X} = ([X_1], [X_2], \dots, [X_n])$ with $X_i \in \mathbb{C}^r$ ($i = 1, 2, \dots, n$).

Now we observe i.i.d. complex r-variate random vectors $X_1, X_2, \dots, X_n$. Here we assume that each X_i ($i = 1, 2, \dots, n$) follows an r-variate complex normal distribution, the density function of which is given by

$$
\begin{aligned}
f_X(\boldsymbol{x}) &= \frac{1}{(2\pi)^r \operatorname{Det} \boldsymbol{\Lambda}(\boldsymbol{\Sigma})^{1/2}} \exp\Big(-\frac{1}{2}[\boldsymbol{x}]' \boldsymbol{\Lambda}(\boldsymbol{\Sigma})^{-1}[\boldsymbol{x}]\Big) \\
&= \frac{1}{\pi^r \det(2\boldsymbol{\Sigma})} \exp\Big(-\boldsymbol{x}^*(2\boldsymbol{\Sigma})^{-1}\boldsymbol{x}\Big)
\end{aligned}
$$

where $\boldsymbol{\Sigma} \in \mathbb{C}_+^{r\times r}$ and $\boldsymbol{x}^*$ stands for the transpose complex conjugate of $\boldsymbol{x} \in \mathbb{C}^r$. We consider the problem of estimating σ based on $X_1, X_2, \ldots, X_n$ under the loss function (9). Since $N = n$ and $g = 2$, the density function of $\boldsymbol{W} = Q([X_1], [X_2], \ldots, [X_n])$ is given by

$$
\frac{\det(\boldsymbol{W})^{n-r} \exp\{-\operatorname{tr}(\boldsymbol{W} \circ (2\boldsymbol{\Sigma})^{-1})\}}{\det(2\boldsymbol{\Sigma})^n \pi^{(r(r-1)/2)} \prod_{j=1}^r \Gamma(n+1-j)}, \qquad \boldsymbol{W} \in \mathbb{C}_+^{r\times r},
$$

which is a special case of (8) for $v = r^2$ and $g = 2$. Here we take the Lebesgue measure on $\mathbb{C}_+^{r\times r}$ as $d\boldsymbol{W} = \prod_{1\le i\le j\le r} da_{ij} \prod_{1\le i<j\le r} db_{ij}$ for $\boldsymbol{W} = (a_{ij} + \boldsymbol{i} b_{ij})$.

Since $p = 2r$, the maximum likelihood estimator for $\boldsymbol{\Sigma}$ is given by $\widehat{\boldsymbol{\Sigma}}_{mle} = (2n)^{-1}\boldsymbol{W}$, the risk of which is given by

$$
\mathscr{R}(\widehat{\boldsymbol{\Sigma}}_{mle}, \boldsymbol{\Sigma}) = r \log n + r \log 2 - \sum_{j=1}^{r} \mathbb{E}[\log u_j^2]
$$

with $\mathcal{L}(u_j^2) = \chi^2_{2(n+1-j)}$ $(j = 1, 2, \ldots, r)$. Next we consider estimators satisfying $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^*) = \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{A}^*$ for all lower-triangular matrices $\boldsymbol{A}$ with positive diagonal elements in order to obtain a minimax risk. Standard arguments such as [27, Section 4.3] and [10, Section 6.3] show that lower-triangular invariant estimators have the form $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{T}\boldsymbol{\Delta}\boldsymbol{T}^*$ where $\boldsymbol{T}$ is a lower-triangular matrix such that $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}^*$, and $\boldsymbol{\Delta}$ is a positive-definite Hermitian constant matrix. Then from Proposition 9 we see that the risk function for $\boldsymbol{T}\boldsymbol{\Delta}\boldsymbol{T}^*$ is minimized when $\boldsymbol{\Delta}$ is a diagonal matrix with the j-th diagonal element $\{2(n+r-2j+1)\}^{-1}$ $(j = 1, 2, \ldots, r)$, and that the minimax risk is given by

$$
\sum_{j=1}^{r} \{\log(n+r-2j+1) - \mathbb{E}[\log u_j^2]\} + r \log 2,
$$

where $\mathcal{L}(u_j^2) = \chi^2_{2(n+1-j)} (j = 1, 2, \ldots, r)$. Now consider estimators of the form

$$
\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{U}^*, \tag{20}
$$

where $\boldsymbol{W} = \boldsymbol{U}\,\mathbf{diag}\,(\boldsymbol{a})\boldsymbol{U}^*$ with a unitary matrix $\boldsymbol{U}$ and $\boldsymbol{a} = (a_1, a_2, \ldots, a_r)$ is a vector of ordered eigenvalues of $\boldsymbol{W}$ with $a_1 \ge a_2 \ge \cdots \ge a_r$, and $\boldsymbol{\Phi} = \mathbf{diag}\,(\varphi_1(\boldsymbol{a}), \ldots, \varphi_r(\boldsymbol{a}))$ in which $\varphi_j(\boldsymbol{a})$'s $(j = 1, 2, \ldots, r)$ are real-valued functions. For a row vector $\boldsymbol{x} = (x_1, \ldots, x_r)$ we denote by $\mathbf{diag}\,(\boldsymbol{x})$ a diagonal matrix with the j-th element x_j $(j = 1, 2, \ldots, r)$. From Theorem 11 a sufficient condition for the estimators (20) to be minimax is given by

$$\sum_{j=1}^{r}\left\{2\frac{\partial\varphi}{\partial a_j}+2(n-r)\frac{\varphi_j}{a_j}+4\sum_{i=j+1}^{r}\frac{\varphi_i-\varphi_j}{a_i-a_j}-\log\varphi_j\right\}\leq-\sum_{j=1}^{r}\log\delta_j.$$

where $\delta_j^{-1}=2(n+r-2j+1)$ $(j=1,\,2,\ldots,\,r)$. Applying Proposition 12 or the inequality above, we see that the estimator $\widehat{\boldsymbol{\Sigma}}_m=\boldsymbol{U}\mathrm{diag}(\delta_1 a_1,\,\ldots\,,\,\delta_r a_r)\boldsymbol{U}^*$ is minimax. Furthermore, we can see that Stein's rough estimator for $\boldsymbol{\Sigma}$ is obtained as $\widehat{\boldsymbol{\Sigma}}_{st}=\boldsymbol{U}\mathbf{diag}\,(\phi_1^{(st)}a_1,\,\ldots\,,\,\phi_r^{(st)}a_r)\boldsymbol{U}^*$, where

$$\phi_j^{(st)}=1\Bigg/\left(2n-2r+4\sum_{\ell\neq j}\frac{a_j}{a_j-a_\ell}\right).$$

The estimator $\widehat{\boldsymbol{\Sigma}}_{st}$ extends Stein's rough estimator for a real normal covariance matrix to that for a covariance matrix of the multivariate complex normal distribution. Then a modification to the estimators $\phi_j^{(st)}$'s is obtained from an isotonizing procedure described in [21].

4.2 Quaternion Normal distributions

For any $\boldsymbol{H}\in\mathbb{H}^{q\times r}$ and $\boldsymbol{h}\in\mathbb{H}^r$, write $\boldsymbol{H}=\boldsymbol{A}+\boldsymbol{i}\boldsymbol{B}+\boldsymbol{j}\boldsymbol{C}+\boldsymbol{k}\boldsymbol{D}$ and $\boldsymbol{h}=\boldsymbol{a}+\boldsymbol{i}\boldsymbol{b}+\boldsymbol{j}\boldsymbol{c}+\boldsymbol{k}\boldsymbol{d}$, where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}\in\mathbb{R}^{q\times r}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}\in\mathbb{R}^r$, and $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are the quaternion units satisfying

$$\boldsymbol{i}^2=\boldsymbol{j}^2=\boldsymbol{k}^2=-1,\quad \boldsymbol{i}\boldsymbol{j}=\boldsymbol{k}=-\boldsymbol{j}\boldsymbol{i},\quad \boldsymbol{j}\boldsymbol{k}=\boldsymbol{i}=-\boldsymbol{k}\boldsymbol{j},\quad \boldsymbol{k}\boldsymbol{i}=\boldsymbol{j}=-\boldsymbol{i}\boldsymbol{k}.$$

We denote by $\boldsymbol{h}^*=\boldsymbol{a}'-\boldsymbol{i}\boldsymbol{b}'-\boldsymbol{j}\boldsymbol{c}'-\boldsymbol{k}\boldsymbol{d}'$ the transpose quaternion conjugate of $\boldsymbol{h}$. Then we define the partitioned vector $[\boldsymbol{h}]$ and the matrix $\{\boldsymbol{H}\}$ as

$$[\boldsymbol{h}]=\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ \boldsymbol{c}\\ \boldsymbol{d}\end{pmatrix},\qquad \{\boldsymbol{H}\}=\begin{pmatrix}\boldsymbol{A}&-\boldsymbol{B}&-\boldsymbol{C}&-\boldsymbol{D}\\ \boldsymbol{B}&\boldsymbol{A}&-\boldsymbol{D}&\boldsymbol{C}\\ \boldsymbol{C}&\boldsymbol{D}&\boldsymbol{A}&-\boldsymbol{B}\\ \boldsymbol{D}&-\boldsymbol{C}&\boldsymbol{B}&\boldsymbol{A}\end{pmatrix}.$$

Now let $\mathbb{H}_H^{r\times r}$ be the vector space of Hermitian $r\times r$ matrices with quaternion entries with the Jordan product given by (19). Then it is known that $\mathbb{H}_H^{r\times r}$ with the multiplication $\circ$ is a Euclidean simple Jordan algebra with the rank r and the Peirce invariant $g=4$. Furthermore, we define the representation of $\boldsymbol{\Lambda}$ of $\mathbb{H}_H^{r\times r}$ in $\mathbb{R}^{4r}$ by $\boldsymbol{\Lambda}(\boldsymbol{H})[\boldsymbol{x}]=[\boldsymbol{H}\boldsymbol{x}]$ for any $\boldsymbol{x}\in\mathbb{H}^r$ and $\boldsymbol{H}\in\mathbb{H}_H^{r\times r}$. From [23, page 380] and Lemma 3(ii), we have

$$\mathrm{Det}\,\{\boldsymbol{H}\}=(\det\boldsymbol{H})^4\qquad\text{and}\qquad\mathrm{Tr}\{\boldsymbol{H}\}=4\mathrm{tr}\,\boldsymbol{H}\tag{21}$$

for any $\boldsymbol{H} \in \mathbb{H}_H^{r\times r}$. For $\mathbb{H}_H^{r\times r}$, we fix a Jordan frame $c_1, c_2, \dots, c_r$ which are defined in the previous section. Then an element in the general linear group, the space of $r \times r$ nonsingular matrices with quaternion entries, acts on an element $\boldsymbol{X} \in \mathbb{H}_H^{r\times r}$ in the same way as that in complex case.

Assume that $\boldsymbol{\Sigma}$ belongs to $\mathbb{H}_+^{r\times r}$, the set of all $r\times r$ Hermitian positive definite matrices with quaternion entries, and that we observe i.i.d. random vectors $X_1, X_2, \dots, X_n$ with quaternion entries. Furthermore we assume that each $[X_i]$ $(i = 1, 2, \dots, r)$ follows the $4r$-variate real normal distribution with zero-mean and a covariance matrix $\boldsymbol{\Lambda}(\boldsymbol{\Sigma})$ with quaternion structure, i.e.,

$$f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{2r}\mathrm{Det}\,\boldsymbol{\Lambda}(\boldsymbol{\Sigma})^{1/2}}\exp\left\{-\frac{1}{2}[\boldsymbol{x}]'\boldsymbol{\Lambda}(\boldsymbol{\Sigma})^{-1}[\boldsymbol{x}]\right\}.$$

From (6) and (21) we have $\boldsymbol{W} = Q([X_1], [X_2], \dots, [X_n]) = \sum_{j=1}^n X_i X_i^*/4$. Then, from Proposition 4 with $N = 2n$ and $g = 4$, we can see that the density function of $\boldsymbol{W}$ is given by

$$\frac{2^{2nr}\det(\boldsymbol{W})^{2n-2r+1}\exp\{-2\mathrm{tr}\,(\boldsymbol{W}\circ(4\boldsymbol{\Sigma})^{-1})\}}{\det(4\boldsymbol{\Sigma})^{2n}\pi^{r(r-1)}\Pi_{j=1}^r\Gamma(2(n+1-j))}, \qquad \boldsymbol{W} \in \mathbb{H}_+^{r\times r},$$

which is a special case of (8) for $v = r(2r-1)$ and $g = 4$. Here we take the Lebesgue measure on $\mathbb{H}_+^{r\times r}$ as $d\boldsymbol{W} = \prod_{1\le i\le j\le r} dw_{ij}^{(1)} \prod_{k=2}^4 \prod_{1\le i<j\le r} dw_{ij}^{(k)}$ for $\boldsymbol{W} = (w_{ij}^{(1)} + \boldsymbol{i}w_{ij}^{(2)} + \boldsymbol{j}w_{ij}^{(3)} + \boldsymbol{k}w_{ij}^{(4)})$.

The maximum likelihood estimator for $\boldsymbol{\Sigma}$, $\widehat{\boldsymbol{\Sigma}}_{mle} = \boldsymbol{W}/(4n)$, has its risk given by

$$\mathscr{R}(\widehat{\boldsymbol{\Sigma}}_{mle}, \boldsymbol{\Sigma}) = r\log n + 2r\log 2 - \sum_{j=1}^r \mathbb{E}[\log u_j^2],$$

where $\mathcal{L}(u_j^2) = \chi^2_{4(n+1-j)}$ $(j = 1, 2, \dots, r)$. Consider a class of lower-triangular invariant estimators having the form $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{T\Delta T}^*$ where $\boldsymbol{T}$ is a lower-triangular matrix such that $\boldsymbol{W} = \boldsymbol{TT}^*$, and $\boldsymbol{\Delta}$ is a positive-definite Hermitian constant matrix. Then, from Proposition 9, we see that the risk function for $\boldsymbol{T\Delta T}^*$ is minimized when $\boldsymbol{\Delta}$ is a diagonal matrix with the j-th diagonal element $\{4(n+r-2j+1)\}^{-1}$ $(j = 1, 2, \dots, r)$, and that the minimax risk is given by

$$\sum_{j=1}^r\{\log(n+r-2j+1) - \mathbb{E}[\log u_j^2]\} + 2r\log 2,$$

where $\mathcal{L}(u_j^2) = \chi^2_{4(n+1-j)}(j = 1, 2, \dots, r)$. Next, using diagonalization of Hermitian matrix with quaternion entries (see, for example, [26, Section 8.9]), we put $\boldsymbol{W} = \boldsymbol{U}\mathbf{diag}\,(a_1, \dots, a_r)\boldsymbol{U}^*$, where $\boldsymbol{U}$ is a symplectic matrix such that $\boldsymbol{UU}^* = \boldsymbol{I}_r$. Without loss of generality, we can assume that $a_1, \dots, a_r$ are positive real numbers such that $a_1 \ge \dots \ge a_r$. We apply Proposition 12 to see that the estimator $\widehat{\boldsymbol{\Sigma}}_\delta = \boldsymbol{U}\mathbf{diag}\,(\delta_1 a_1, \dots, \delta_r a_r)\boldsymbol{U}^*$ is minimax in the case that $\delta_j = \{4(n+r-2j+1)\}^{-1}$ $(j = 1, 2, \dots, r)$. Furthermore, we can see that

Stein's rough estimator for $\mathbf{\Sigma}$ is obtained as $\widehat{\mathbf{\Sigma}}_{st} = \boldsymbol{U}\mathbf{diag}\,(\phi_1^{(st)}a_1,\, \ldots\, ,\, \phi_r^{(st)}a_r)\boldsymbol{U}^*$, where

$$\phi_j^{(st)} = 1\bigg/\left(4n - 4r + 2 + 8\sum_{\ell \neq j} \frac{a_j}{a_j - a_\ell}\right).$$

The estimator $\widehat{\mathbf{\Sigma}}_{st}$ for positive-definite Hermitian matrix with quaternion entries extends Stein's rough estimator. Then a modification to the estimators $\phi_j^{(st)}$'s is obtained from an isotonizing procedure described in [21].

4.3 Normal distributions indexed by the Lorentz cones

Recall that $\mathcal{V} = \mathbb{R} \times \mathcal{W}$ is a Jordan algebra associated with a symmetric bilinear form B on $\mathcal{W}$, a real vector space of dimension $v-1$ $(v \geq 3)$, and that a Jordan multiplication is $(a_1,\, a_2)(b_1,\, b_2) = (a_1 b_1 + B(a_2|\, b_2),\, a_1 b_2 + b_1 a_2)$ for $(a_1,\, a_2),\, (b_1,\, b_2) \in \mathbb{R}\times\mathcal{W}$. We denote by $(\mathbb{R}\times\mathcal{W})_+$ the Lorentz cone associated to $\mathbb{R} \times \mathcal{W}$, i.e., $(\mathbb{R} \times \mathcal{W})_+ = \{(x_1,\, x_2) \in (\mathbb{R} \times \mathcal{W}) : x_1^2 - B(x_2|\, x_2) > 0\}$. It is well known that

$$\mathrm{tr}(x) = 2x_1 \qquad \text{and} \qquad \det(x) = x_1^2 - B(x_2|\, x_2) \tag{22}$$

for $x = (x_1,\, x_2) \in \mathbb{R} \times \mathcal{W}$ and that $x^{-1} = (x_1,\, -x_2)/\det(x)$ if $\det(x) \neq 0$. Consider a p-variate multivariate real normal distribution (4) which is indexed by $\sigma \in (\mathbb{R} \times \mathcal{W})_+$. Observe that $\mathbf{\Lambda} : \mathbb{R} \times \mathcal{W} \mapsto \mathbb{R}_H^{p\times p}$ in (4) is a one-to-one Jordan algebra homomorphism, i.e., Λ is a one-to-one linear mapping such that $\mathbf{\Lambda}(xy) = (1/2)(\mathbf{\Lambda}(x)\mathbf{\Lambda}(y) + \mathbf{\Lambda}(y)\mathbf{\Lambda}(x))$ for $x,\, y \in \mathbb{R} \times \mathcal{W}$. Recall that $\mathbb{R}_H^{p\times p}$ is the space of $p \times p$ symmetric matrices.

Let $X_1,\, X_2,\, \ldots\, ,\, X_n$ be a random sample from a normal distribution (4) with $\sigma = (\sigma_1,\, \sigma_2) \in (\mathbb{R} \times \mathcal{W})_+$, and consider the problem of estimating σ based on $\boldsymbol{X} = (X_1,\, X_2,\, \ldots\, ,\, X_n)$. We define a Wishart random variable $w = (w_1,\, w_2)$ in the closure of $(\mathbb{R} \times \mathcal{W})_+$ as $\mathrm{Tr}(\boldsymbol{x}\boldsymbol{x}'\mathbf{\Lambda}(y)) = (y|\, w)$ for any $y = (y_1,\, y_2)$ in $\mathbb{R} \times \mathcal{W}$. From this we have

$$y_1 \mathrm{Tr}\,(\boldsymbol{x}\boldsymbol{x}') + \mathrm{Tr}\,(\boldsymbol{x}\boldsymbol{x}'\mathbf{\Lambda}(0,\, y_2)) = 2y_1 w_1 + 2B(y_2|\, w_2) \tag{23}$$

for any $y = (y_1,\, y_2)$ in $\mathbb{R} \times \mathcal{W}$. Using Lemma 1(i) and (23), we can see that, for $\widehat{\sigma}_{mle} = (\widehat{\sigma}_1,\, \widehat{\sigma}_2)$,

$$\widehat{\sigma}_1 = \frac{1}{np}\mathrm{Tr}\,(\boldsymbol{x}\boldsymbol{x}') \qquad \text{and} \qquad B(y_2|\, \widehat{\sigma}_2) = \frac{1}{np}\mathrm{Tr}\,(\boldsymbol{x}\boldsymbol{x}'\mathbf{\Lambda}(0,\, y_2))$$

for any $y_2 \in \mathcal{W}$. These equations appear in [35]. Therefore, the maximum likelihood estimator for σ is given by $\widehat{\sigma}_{mle} = (2/np)w$. From Corollary 7 with $N = np/4$, we can see that the maximum likelihood estimator $\widehat{\sigma}$ is unbiased for σ.

To describe the estimator given in Proportion 9, we write a Jordan frame as $c_1 = (1/2)(1, h)$ and $c_2 = (1/2)(1, -h)$ with some $h \in \mathcal{W}$, satisfying $B(h|\, h) = 1$. Then note that $c_1^2 = c_1$, $c_2^2 = c_2$, $c_1 c_2 = 0$, and $c_1 + c_2 = (1, 0) = e$. If $x = (x_1, x_2) \in (\mathbb{R} \times \mathcal{W})_+$, and if we set $a_1 = x_1 + \|x_2\|_B$ and $a_2 = x_1 - \|x_2\|_B$ where $\|x_2\|_B = \sqrt{B(x_2|\, x_2)}$, then we have that $\mathrm{tr}(x) = a_1 + a_2$ and $\det(x) = a_1 a_2$. Hence a_1, a_2 $(a_1 \geq a_2)$ are ordered eigenvalues of x with respect to the Jordan frame c_1, c_2. Furthermore, recall that $t(u)$ is given by (3) and that $u = u_1 c_1 + u_2 c_2 + u_{12}$ for positive constants u_1, u_2 and $u_{12} \in \mathcal{V}_{12}$. If $u_{12} = (u_{12}^{(1)}, u_{12}^{(2)}) \in \mathbb{R} \times \mathcal{W}$, then the condition $u_{12} \in \mathcal{V}_{12}$ implies that $u_{12}^{(1)} = 0$ and that $B(u_{12}^{(2)}|\, h) = 0$. Furthermore, we use Lemma 1(vii) and (viii) to see that

$$t(u)(\alpha c_1 + \beta c_2) = \alpha u_1^2 c_1 \oplus \alpha u_1 u_{12} \oplus \left(\frac{\alpha}{2}\|u_{12}\|^2 + \beta u_2^2\right) c_2, \tag{24}$$

for positive constants α, β. From (24) and $B(u_{12}^{(2)}|\, h) = 0$, we may obtain that, if $t(u)e = (w_1, w_2) \in (\mathbb{R} \times \mathcal{W})_+$, then

$$\begin{cases} w_1 = \frac{1}{2}u_1^2 + \frac{1}{2}u_2^2 + \frac{1}{4}\|u_{12}\|^2, \\ B(w_2|\, h) = \frac{1}{2}u_1^2 - \frac{1}{2}u_2^2 - \frac{1}{4}\|u_{12}\|^2, \\ w_2 = \left(\frac{1}{2}u_1^2 - \frac{1}{2}u_2^2 - \frac{1}{4}\|u_{12}\|^2\right) h + u_1 u_{12}^{(2)}, \\ \det(w_1, w_2) = u_1^2 u_2^2. \end{cases}$$

Hence we have

$$u_1 = \sqrt{w_1 + B(w_2|\, h)}, \qquad u_2 = \sqrt{\frac{w_1^2 - \|w_2\|_B^2}{w_1 + B(w_2|\, h)}},$$
$$u_{12}^{(2)} = \frac{1}{\sqrt{w_1 + B(w_2|\, h)}}\Big(w_2 - B(w_2|\, h)h\Big).$$

Again we use (24) to see that the estimator given in Proposition 9 becomes $t(u)(\delta_1 c_1 + \delta_2 c_2) = (x_1, x_2) \in (\mathbb{R} \times \mathcal{W})_+$, where

$$x_1 = \delta_1 \frac{w_1^2 + 2w_1 B(w_2|\, h) + \|w_2\|_B^2}{2(w_1 + B(w_2|\, h))} + \delta_2 \frac{w_1^2 - \|w_2\|_B^2}{2(w_1 + B(w_2|\, h))},$$
$$x_2 = \delta_1 w_2 + (\delta_1 - \delta_2)\frac{w_1^2 - \|w_2\|_B^2}{2(w_1 + B(w_2|\, h))} h,$$
$$\delta_1^{-1} = \frac{np}{2} + (v - 2), \qquad \delta_2^{-1} = \frac{np}{2} - (v - 2). \tag{25}$$

Next, we need following lemma to describe the singular value decomposition of an element in Ω in order to derive orthogonally invariant minimax estimators in the setup of this section.

Lemma 13 *For any $w = (w_1, w_2) \in (\mathbb{R} \times \mathcal{W})_+$, set*

$$z = \frac{1}{\sqrt{4w_1^2 + 2\|w_2\|_B B(w_2|\, h) - 2\|w_2\|_B^2}}(2w_1,\, w_2 - \|w_2\|_B h). \tag{26}$$

Then we have $w = P(z)(a_1c_1 + a_2c_2)$, where $a_1 = w_1 + \|w_2\|_B$ and $a_2 = w_1 - \|w_2\|_B$. Furthermore, we have $\mathrm{Det}(P(z)) = 1$.

Proof. The first assertion can be obtained from a straightforward application of [13, Corollary 1]. Then, using Lemma 1 (ii) and noting that $\det z = 1$, we can complete the proof. □

Recall that $w = (w_1, w_2)$ is an element in $(\mathbb{R}\times\mathcal{W})_+$ such that $\mathrm{Tr}(\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Lambda}(y)) = (y|\, w)$ for any y in $\mathbb{R} \times \mathcal{W}$. We use Lemma 13 to decompose the element w as

$$w = P(z)\{\frac{w_1 + \|w_2\|_B}{2}(1,\, h) + \frac{w_1 - \|w_2\|_B}{2}(1,\, -h)\} = P(z)(w_1,\, \|w_2\|_B h),$$

where z is given by (26). From Proposition 12 we can see that our proposed estimator is given by

$$\begin{aligned}\widehat{\sigma}_m &= P(z)(\delta_1 a_1 c_1 + \delta_2 a_2 c_2)\\ &= P(z)\Big\{\delta_1\frac{w_1 + \|w_2\|_B}{2}(1,\, h) + \delta_2\frac{w_1 - \|w_2\|_B}{2}(1,\, -h)\Big\},\end{aligned}$$

where δ_1, δ_2 are given by (25) and that it is minimax. Furthermore, Stein's rough estimator is given by

$$\widehat{\sigma}_{st} = P(z)\Big\{\phi_1^{(st)}\frac{w_1 + \|w_2\|_B}{2}(1,\, h) + \phi_2^{(st)}\frac{w_1 - \|w_2\|_B}{2}(1,\, -h)\Big\},$$

where

$$\begin{aligned}\phi_1^{(st)} &= 1\Bigg/\left(\frac{np}{2} - v + \frac{(v-2)(w_1 + \|w_2\|_B)}{\|w_2\|_B}\right),\\ \phi_2^{(st)} &= 1\Bigg/\left(\frac{np}{2} - v - \frac{(v-2)(w_1 - \|w_2\|_B)}{\|w_2\|_B}\right).\end{aligned}$$

Then a modification to the estimators $\phi_1^{(st)}$ and $\phi_2^{(st)}$ is obtained from an isotonizing procedure described in [21].

Besides two estimators stated above, we can construct explicit form of orthogonally invariant estimators using approaches due to [28,34]. To describe these two approaches in a unified manner, let γ be a positive constant. We again decompose w into $P(z)(a_1c_1 + a_2c_2)$ where $P(z)$, a_1, and a_2 are defined as in Lemma 13, and we consider a class of estimators for σ, being of the form

$$\widehat{\sigma}_{pt} = P(z)\{\phi_1^{(pt)} a_1 c_1 + \phi_2^{(pt)} a_2 c_2\}, \tag{27}$$

where

$$\phi_1^{(pt)} = \left(\frac{a_1^\gamma}{a_1^\gamma + a_2^\gamma}\delta_1 + \frac{a_2^\gamma}{a_1^\gamma + a_2^\gamma}\delta_2\right) \quad \text{and} \quad \phi_1^{(pt)} = \left(\frac{a_2^\gamma}{a_1^\gamma + a_2^\gamma}\delta_1 + \frac{a_1^\gamma}{a_1^\gamma + a_2^\gamma}\delta_2\right).$$

If $\gamma = 1$ then the estimator (27) corresponds to the estimator for the normal covariance matrix given by [28] while it corresponds to the estimator for the normal covariance matrix given by [34] if $\gamma = 1/2$.

Proposition 14 *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independently and identically distributed as (4) with $\kappa = 1/2$ for some σ in $(\mathbb{R} \times \mathcal{W})_+$ and assume that w is an element in the closure of $(\mathbb{R} \times \mathcal{W})_+$ such that $\operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Lambda}(y)) = (y|\, w)$ for any y in $\mathbb{R} \times \mathcal{W}$. Then the estimators given by (27) are minimax for any positive constant γ.*

Proof. We apply Theorem 11 with $\varphi_j = \phi_j^{(pt)} a_j$ $(j = 1, 2)$, $r = 2$, and $\kappa = g = 1/2$. From straightforward calculation we have

$$\sum_{j=1}^{2} \frac{\partial \varphi_j}{\partial a_j} = \delta_1 + \delta_2 + \frac{2\gamma(a_1a_2)^\gamma}{(a_1^\gamma + a_2^\gamma)^2}(\delta_1 - \delta_2) \leq \delta_1 + \delta_2, \tag{28}$$

$$\frac{\varphi_1 - \varphi_2}{a_1 - a_2} = \delta_1 + \frac{a_1a_2(a_1^{\gamma-1} - a_2^{\gamma-1})}{(a_1 - a_2)(a_1^\gamma + a_2^\gamma)}(\delta_1 - \delta_2) \leq \delta_1. \tag{29}$$

Furthermore, from the convexity of the logarithmic function and Jensen's inequality we have

$$\log\det(\widehat{\sigma}_{pt}) = \log(\phi_1^{(pt)}\phi_2^{(pt)}) + \log(a_1a_2) \geq \sum_{j=1}^{2}\{\log a_j + \log\delta_j\}. \tag{30}$$

Putting (28)-(30) into (16) and using Corollary **??**, we have

$$\begin{aligned}\mathscr{R}(\widehat{\sigma}_{pt}, \sigma) &\leq 2(\delta_1 + \delta_2) + \left(\frac{np}{2} - p\right)(\delta_1 + \delta_2) + 2(p-2)\delta_1 \\ &\qquad - \log(\delta_1\delta_2) - \mathbb{E}[\log(a_1a_2)] + \log\det\sigma - 2 \\ &= -\sum_{j=1}^{2}\{\mathbb{E}[\log u_j^2] + \log\delta_j\},\end{aligned}$$

where $\mathcal{L}(u_j^2) = \chi^2_{(np/2)-(v-2)(j-1)}$ $(j = 1, 2)$. This completes the proof. □

Acknowledgements

The author is very grateful to the referees for their careful evaluation and very detailed comments and suggestions that have greatly improved the presentation of the paper.

References

[1] H.H. Andersen, M. Højbjerre, D. Sørensen, P.S. Eriksen, Linear and Graphical Models, Springer-Verlag, New York (1995).

[2] S.A. Andersson, Invariant normal models, Ann. Statist. 3 (1975) 132-154.

[3] S.A. Andersson, M.D. Perlman, Two testing problems relating the real and complex multivariate normal distributions, J. Multivariate Anal. 15 (1984) 21-51.

[4] S.A. Andersson, M.D. Perlman, Lattice models for conditional independence in a multivariate normal distribution, Ann. Statist. 21 (1993) 1318-1358.

[5] S.A. Andersson, G.G. Wojnar, The Wishart distributions on homogeneous cones, J. Theoret. Prob. 16 (2004) 781-818.

[6] J. Bonder, P. Milnes, Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups, Z. Wahrsch. Verw. Gebiete 57 (1981) 103-128.

[7] M. Casalis, G. Letac, The Lukas-Olkin-Rubin characterization of Wishart distributions on symmetric cones, Ann. Statist. 24 (1996) 763-786.

[8] G. Consonni, P. Veronese, Enriched conjugate and reference priors for the Wishart family on symmetric cones, Ann. Statist. 31 (2003) 1491-1516.

[9] D.K. Dey, C. Srinivasan, Estimation of a covariance matrix under Stein's loss, Ann. Statist. 13 (1985) 1581-1591.

[10] M.L. Eaton, Group Invariance Application in Statistics, Regional conference series in Probability and Statistics Vol. 1, Institute of Mathematical Statistics (1989).

[11] J. Faraut, K. Korányi, Analysis on Symmetric Cones, Oxford Science Publications (1994).

[12] L. Faybusovich, Euclidean Jordan algebras and interior-point algorithms, Positivity 1 (1997) 331-357.

[13] L. Faybusovich, T. Tsuchiya, Primal-dual algorithms and infinite-dimensional Jordan algebras of finite rank, Math. Program Ser. B 97 (2003) 471-493.

[14] N.R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction), Ann. Math. Statist. 34 (1963) 152-176.

[15] P. Graczyk, G. Letac, H. Massam, The complex Wishart distribution and symmetric group, Ann. Statist. 31 (2003) 287-309.

[16] L.R. Haff, The variational form of certain Bayes estimators, Ann. Statist. 19 (1992) 1163-1190.

[17] W. James, C. Stein, Estimation with quadratic loss, in Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1 361-380, Univ. California Press (1961).

[18] C.G. Khatri, Classical statistical analysis based on a certain multivariate complex Gaussian distribution, Ann. Math. Statist. 36 (1965) 98-114.

[19] Y., Konno, Inadmissibility of the maximum likelihood estimator of *J. Multivariate Anal.* **79** (2001) 33-51.

[20] G. Letac, H. Massam, Representation of the Wishart distributions, in Contemporary Mathematics, **261** (2000) 121-142.

[21] S.P. Lin, M.D. Perlman, A Monte Carlo comparison of four estimators for a covariance matrix, in Multivariate Analysis VI (P.R. Krishnaiah, ed.) 411-426, North-Holland, Amsterdam (1985).

[22] D. Maiward, D. Kraus, Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices, Radar, Sonar and Navigation, IEEE Proceedings 147 2000 162 - 168.

[23] H. Massam, An Exact decomposition theorem and a unified view of some related distributions for a class of exponential transformation models on symmetric cones, Ann. Statist. 32 (1994) 389-394.

[24] H. Massam, E. Neher, On transformations and determinants of Wishart variables on symmetric cones, J. Theoret. Probat. 10 (1997) 867-902.

[25] H. Massam, E. Neher, Estimation and testing for lattice conditional independence models of Euclidean Jordan algebras, Ann. Statist. 26 (1998) 1051-1082.

[26] M.L. Mehta, Matrix Theory, Les Editions de Phsique, 91944 Les Ulis Cedex, France (1989).

[27] R.J. Muirhead, Aspects of Multivariate Statistical Analysis, John Wiley & Sons, Inc (1982).

[28] F. Perron, Minimax estimators of a covariance matrix, J. Multivariate Anal. 43 (1992) 6-28.

[29] P. Shaman, The inverted complex Wishart distribution and its application to spectral estimation, J. Multivariate Anal. 10 (1980) 51-59.

[30] Y. Sheena, Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix, J. Japan Statist. Soc. 25 (1995) 35-48.

[31] Y. Sheena, A. Takemura, Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss, J. Multivariate Anal. 41 (1992) 117-131.

[32] C. Stein, Lectures on the theory of estimation of many parameters, in Studies in the Statistical Theory of Estimation I(I. A. Ibragimov and M. S. Nikulin, eds.) (1977).

[33] L. Svensson, M. Lundberg, Estimating complex covariance matrix, Signals, Systems and Computers, Conference Record of the Thirty-Eighth Asilomar Conference on Volume 2 (2004) 7-10, 2151 - 2154.

[34] A. Takemura, An orthogonally invariant minimax estimators of the covariance matrix of a multivariate normal populations, Tsukuba J. Math. 8 (1984) 365-376.

[35] S. Tolver Jensen, Covariance hypotheses which are linear in both the covariance and the inverse covariance, Ann. Statist. 16 (1988) 302-322.

5 Appendix

Derivation of the representation of $t(u)(\delta_1 c_1 + \delta_2 c_2)$ where $\delta_1 > 0$, $\delta_2 > 0$.

Recall that $t(u) = P(b_1)\tau_{c_1}(u^{(1)})P(b_2)$ where $b_1 = u_1 c_1 + c_2$, $b_2 = c_1 + u_2 c_2$, and $u^{(1)} = u_{12}$. First we compute $P(b_2)(\delta_1 c_1 + \delta_2 c_2)$ as

$$\begin{aligned} P(b_2)(\delta_1 c_1 + \delta_2 c_2) &= \{2L(b_2)^2 - L(b_2)\}(\delta_1 c_1 + \delta_2 c_2) \\ &= 2L(b_2)(c_1 + u_2 c_2)(\delta_1 c_1 + \delta_2 c_2) - (c_1 + u_2^2 c_2)(\delta_1 c_1 + \delta_2 c_2) \\ &= 2(c_1 + u_2 c_2)(\delta_1 c_1 + \delta_2 u_2 c_2) - (\delta_1 c_1 + \delta_2 u_2^2 c_2) \\ &= 2(\delta_1 c_1 + \delta_2 u_2^2 c_2) - (\delta_1 c_1 + \delta_2 u_2^2 c_2) \\ &= \delta_1 c_1 + \delta_2 u_2^2 c_2. \end{aligned}$$

Next we use Lemma 1(vii) and (viii) in order to compute $\tau_{c_1}(u_{12})(\delta_1 c_1 + \delta_2 u_2^2 c_2)$ as

$$\tau_{c_1}(u_{12})(\delta_1 c_1 + \delta_2 u_2^2 c_2) = \delta_1 c_1 \oplus \delta_1 u_{12} \oplus \{2L(e - c_1)L(u_{12})^2 \delta_1 c_1 + \delta_2 u_2^2 c_2\}.$$

Furthermore we have

$$2L(e - c_1)L(u_{12})^2 \delta_1 c_1 = \delta_1 L(e - c_1)u_{12}^2 = \frac{1}{2}L(e - c_1)\|u_{12}\|^2(c_1 + c_2) = \frac{1}{2}\|u_{12}\|^2 c_2,$$

from which it follows that

$$\tau_{c_1}(u_{12})(\delta_1 c_1 + \delta_2 u_2^2 c_2) = \delta_1 c_1 \oplus \delta_1 u_{12} \oplus \left(\frac{1}{2}\delta_1\|u_{12}\|^2 + \delta_2 u_2^2\right) c_2.$$

Since

$$\begin{aligned} P(b_1)u_{12} &= 2L(b_1)^2 u_{12} - L(b_1^2)u_{12} = 2u_1 L(b_1)(u_1 c_1 + c_2)u_{12} - (u_1^2 c_1 + C_2)u_{12} \\ &= L(b_1)u_{12} - \frac{1}{2}(u_1^2 + 1)u_{12} \\ &= (u_1 + 1)(u_1 c_1 + c_2)u_{12} - \frac{1}{2}(u_1^2 + 1)u_{12} \\ &= \frac{1}{2}(u_1 + 1)^2 u_{12} - \frac{1}{2}(u_1^2 + 1)u_{12} = u_1 u_{12}, \end{aligned}$$

we have

$$\begin{aligned} t(u)(\delta_1 c_1 + \delta_2 c_2) &= \delta_1 P(b_1)c_1 \oplus \delta_1 P(b_1)u_{12} \oplus \left(\frac{1}{2}\delta_1\|u_{12}\|^2 + \delta_2 u_2^2\right) c_2 \\ &= \delta_1 u_1^2 c_1 \oplus \delta_1 u_1 u_{12} \oplus \left(\frac{\delta_1}{2}\|u_{12}\|^2 + \delta_2 u_2^2\right) c_2. \end{aligned}$$

Putting $c_1 = (1/2)(1,\, h)$, $u_{12} = (0,\, u_{12}^{(2)})$ and $c = (1/2)(1,\, -h)$ into the right hand side of the above equation, we have

$$t(u)(\delta_1 c_1 + \delta_2 c_2) = \frac{\delta_1 u_1^2}{2}(1,\, h) + \delta_1 u_1 (0,\, u_{12}^{(2)}) + \frac{1}{2}\left(\frac{\delta_1}{2}\|u_{12}\|^2 + \delta_2 u_{12}^2\right)(1,\, -h)$$
$$= \left(\frac{1}{2}\delta_1 u_1^2 + \frac{\delta_1 \|u_{12}\|^2}{4} + \frac{\delta_2 u_2^2}{2},\, \Big(\frac{\delta_1 u_1}{2} - \frac{\delta_2 u_2^2}{2} - \frac{\delta_1 \|u_{12}\|^2}{4}\Big)h + \delta_1 u_1 u_{12}^{(2)}\right).$$

Furthermore note that

$$\|u_{12}^{(2)}\|_B^2 = \frac{\|w_2\|_B^2 - B^2(w_2|\, h)}{w_1 + B(w_2|\, h)},$$

and that, from (22),

$$\|u_{12}\|^2 = \mathrm{tr}(\, u_{12}^2) = \mathrm{tr}\{\,(0,\, u_{12}^{(2)})(0,\, u_{12}^{(2)})\} = \mathrm{tr}\{\,(B(u_{12}^{(2)}|\, u_{12}^{(2)}),\, 0)\} = 2\|u_{12}^{(2)}\|_B^2,$$

where $t(u)e = (w_1,\, w_2)$. Therefore, we have $t(u)(\delta_1 c_1 + \delta_2 c_2) = (x_1,\, x_2)$, where

$$\begin{aligned}
x_1 &= \frac{\delta_1}{2}\left(u_1^2 + \|u_{12}^{(2)}\|_B^2\right) + \frac{\delta_2 u_2^2}{2}\\
&= \frac{\delta_1}{2}\left(w_1 + B(w_2|\, h) + \frac{\|w_2\|_B^2 - B^2(w_2|\, h)}{w_1 + B(w_2|\, h)}\right) + \frac{\delta_2}{2}\frac{w_1^2 - \|w_2\|_B^2}{w_1 + B(w_2|\, h)}\\
&= \delta_1 \frac{w_1^2 + 2w_1 B(w_2|\, h) + \|w_2\|_B^2}{2(w_1 + B(w_2|\, h)} + \delta_2 \frac{w_1^2 - \|w_2\|_B^2}{2(w_1 + B(w_2|\, h))}
\end{aligned}$$

and

$$\begin{aligned}
x_2 &= \frac{\delta_1}{2}\left(u_1^2 - \|u_{12}^{(2)}\|_B^2\right) h + \delta_1 u_1 u_{12}^{(2)} - \frac{\delta_2}{2} u_2^2 h\\
&= \frac{\delta_1}{2}\left((w_1 + B(w_2|\, h)) - \frac{\|w_2\|_B^2 - B^2(w_2|\, h)}{w_1 + B(w_2|\, h)}\right) h + \delta_1 (w_2 - B(w_2|\, h)h)\\
&\quad - \frac{\delta_2}{2}\frac{w_1^2 - \|w_2\|_B^2}{w_1 + B(w_2|\, h)} h\\
&= \delta_1 w_2 + \frac{\delta_1}{2}\frac{w_1 - \|w_2\|_B^2}{w_1 + B(w_2|\, h)} h - \frac{\delta_2}{2}\frac{w_1^2 - \|w_2\|_B^2}{w_1 + B(w_2|\, h)} h.
\end{aligned}$$