ESTIMATION OF MULTIVARIATE COMPLEX NORMAL COVARIANCE MATRICES UNDER AN INVARIANT QUADRATIC LOSS

YOSHIHIKO KONNO

email:konno@fc.jwu.ac.jp
Japan Women's University, Tokyo 112-8681, Japan

June 6, 2008

The problem of estimating multivariate complex normal covariance matrices is considered under an invariant quadratic loss function. Estimators which dominate the best scalar multiple of the empirical covariance matrix are presented. Improved estimators include the best triangular equivariant estimators and complex analogues of estimators for multivariate real normal covariance matrices by Haff [1980, *Ann. Statist.* 8:586-597].

Keywords: Complex Wishart distributions; Minimax estimators; Stein-Haff identity; Calculus on eigenstructures; Shrinkage estimators; Hunt-Stein theorem.

Mathematics Subject Classification: 62H12; 62F10.

1. INTRODUCTION

The theory of complex random matrices plays a vital role in multidimensional signal processing algorithms[see Maiward and Kraus (1994), Tague and Caldwell (1994), Ratnarajah and Vaillancourt (2005) for examples]. Among others, complex Wishart matrices are important in some applications. In this paper we consider the problem of estimating multivariate complex normal covariance matrices in a decision-theoretic point of view.

Consider p-dimensional complex random column vectors $Z_1, Z_2, \ldots, Z_n$ which are independently and identically distributed as a multivariate complex normal distribution with zero mean vector and a $p \times p$ positive definite Hermitian covariance matrix $\mathbf{\Sigma}$. We denote by $\mathbb{C}N_p(\mathbf{0}, \mathbf{\Sigma})$ this distribution. The probability density function of $\mathbb{C}N_p(\mathbf{0}, \mathbf{\Sigma})$ is

$$f_Z(z) = \frac{1}{\pi^p \mathrm{Det}\,(\mathbf{\Sigma})} \exp(-z^* \mathbf{\Sigma}^{-1} z), \qquad z \in \mathbb{C}^p,$$

where the superscript "*" and Det denote the complex conjugate transpose of a column complex vector or a matrix and the determinant of a square matrix, respectively. See Andersen *et al.* (1995), Goodman (1963), Giri (2004), and Khatri (1965) for basic properties of the multivariate complex normal distributions.

Put $\boldsymbol{W} = \sum_{i=1}^n Z_i Z_i^*$. Then $\boldsymbol{W}$ is positive definite with probability 1, and has the so-called 'complex Wishart distribution' with the dimension p, the degrees of freedom n, and the scale matrix $\mathbf{\Sigma}$. We denote by $\mathbb{C}W_p(\mathbf{\Sigma}, n)$ this distribution. If $n \geq p$, then the the density function of $\boldsymbol{W}$ with respect to the Lebesgue measure on $\mathbb{C}_+^{p\times p}$, the set of $p \times p$ positive definite complex matrices, is given as

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{\mathrm{Det}\,(\boldsymbol{w})^{n-p} \exp\{-\mathrm{Tr}\,(\boldsymbol{w}\mathbf{\Sigma}^{-1})\}}{\mathrm{Det}\,(\mathbf{\Sigma})^p \pi^{p(p-1)/2} \Pi_{i=1}^n \Gamma(n+1-i)}, \qquad \boldsymbol{w} \in \mathbb{C}_+^{p\times p}, \qquad (1.1)$$

where Tr and $\Gamma(\,\cdot\,)$ denote the trace of a matrix and Euler's gamma function, respectively. See Andersen *et al.* (1995), Giri (2004), Khatri (1965), Srivastava (1965) for basic properties of the complex Wishart distributions.

The general linear group $GL_p(\mathbb{C})$, the set of $p \times p$ invertible complex matrices, acts on the sample space and the parameter space as

$$\boldsymbol{W} \mapsto \boldsymbol{g}\boldsymbol{W}\boldsymbol{g}^*, \qquad \boldsymbol{\Sigma} \mapsto \boldsymbol{g}\boldsymbol{\Sigma}\boldsymbol{g}^*,$$

where $\boldsymbol{g}$ is a $p \times p$ nonsingular complex matrix. A loss function L is said to be invariant if L satisfies

$$L(\boldsymbol{g}\widehat{\boldsymbol{\Sigma}}\boldsymbol{g}^*, \boldsymbol{g}\boldsymbol{\Sigma}\boldsymbol{g}^*) = L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}),$$

where $\widehat{\boldsymbol{\Sigma}} := \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})$ is an estimator of $\boldsymbol{\Sigma}$. Invariant loss functions used in the literature are

$$\begin{aligned} L_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2; \\ L_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log \mathrm{Det}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - p; \\ L_3(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) + \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) - 2p. \end{aligned}$$

For multivariate real normal covariance matrix estimation, the loss function L_1 was used by Olkin and Selliah (1977). James and Stein (1961) introduced the loss L_2. Svensson (2004) and Konno (2007b) employed the loss L_2 for the multivariate complex normal covariance estimation, and obtained a counterpart of the results for the multivariate real normal covariance estimation by Stein (1977), Dey and Srinivasan (1985), Haff (1988,1991), and Kubokawa and Srivastava (1999). Consonni and Veronese (2003) and Konno (2007a) generalized some results under the loss L_2 to the Wishart distributions on the symmetric cones. The loss L_3 was considered by Kubokawa and Konno (1990).

In this paper, we consider the problem of estimating $\boldsymbol{\Sigma}$ based on the nonsingular complex Wishart matrices $\boldsymbol{W}$ under the loss function L_1 and give a counterpart of the results for the multivariate real normal covariance matrix estimation due to Olkin and Selliah (1977) and Haff (1980). In Section 2, using Bartlett's decomposition for the complex Wishart distribution, we give the best triangular equivariant estimators which are counterpart of the estimators for the multivariate real normal covariance matrix due to Olkin and Selliah (1977). In Section 3, we state integration by parts formula and calculus on eigenstructures for the complex nonsingular Wishart matrices by Svensson (2004) and Konno (2007b). Real analogues of these techniques have been extensively explored in the literature which includes Haff (1988, 1991), Konno (1992), Loh (1988, 1991), and Stein (1977). Using these techniques we derive an unbiased risk estimate for unitary invariant estimators, from which we show that complex analogues of estimators by Haff (1980) dominate the best scalar multiple of the empirical covariance matrix.

Bold uppercase letters represent matrices. Superscripts "$'$" means the transpose of a matrix. For real numbers $a_1, a_2, \ldots, a_n$, we denote by $\mathbf{Diag}(a_1, a_2, \ldots, a_p)$ a $p \times p$ diagonal matrix with diagonal elements $a_1, a_2, \ldots, a_p$. For a complex number $z \in \mathbb{C}$, we write $z = \mathrm{Re}\, z + \sqrt{-1}\mathrm{Im}\, z$, where $\mathrm{Re}\, z$ and $\mathrm{Im}\, z$ are real numbers. We denote by $\bar{z}$ the complex conjugate of a complex number z, i.e., $\bar{z} = \mathrm{Re}\, z - \sqrt{-1}\mathrm{Im}\, z$. Also we denote $|z|^2 = \bar{z}z$. A continuous function $f : \mathbb{C} \to \mathbb{R}$ is called differentiable on $\mathbb{C}$ if $\partial f/\partial(\mathrm{Re}\, z)$ and $\partial f/\partial(\mathrm{Im}\, z)$ exist on $\mathbb{C}$. For $u, v : \mathbb{C} \to \mathbb{R}$, a function $f = u + \sqrt{-1}v$ is called differentiable if both u and v are differentiable.

2. Bartlett's decomposition and minimax estimators

In this section, we give complex analogues of the James and Stein (1961) estimator for the multivariate real normal covariance matrices[see Eaton (1989) and Muirhead (1982)].

Recall that $\boldsymbol{W}$ follows the complex Wishart distribution $\mathbb{C}W_p(\boldsymbol{\Sigma}, n)$. Decompose $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{Q}\boldsymbol{Q}^*$, where $\boldsymbol{T}$ and $\boldsymbol{Q}$ are $p \times p$ lower and upper triangular matrices with real and positive diagonal elements, respectively. Now consider classes of estimators of the form

$$\widehat{\mathbf{\Sigma}}_T = \boldsymbol{T}\boldsymbol{D}_T\boldsymbol{T}^*; \tag{2.1}$$

$$\widehat{\mathbf{\Sigma}}_Q = \boldsymbol{Q}\boldsymbol{D}_Q\boldsymbol{Q}^*; \tag{2.2}$$

where $\boldsymbol{D}_T$ and $\boldsymbol{D}_Q$ are $p \times p$ diagonal matrices with positive diagonal elements. These estimators $\widehat{\mathbf{\Sigma}}_T$ ($\widehat{\mathbf{\Sigma}}_Q$) are invariant under the transformation of a lower(upper) triangular matrix group which acts transitively on the parameter space, the set of $p \times p$ positive definite Hermitian matrices. Since the loss L_1 is invariant, these estimators have a constant risk. To obtain the best estimators among those of the form (2.1) or (2.2), we use Bartlett's decomposition stated below, which was given by Goodman (1963).

Lemma 2.1. *Assume that a $p \times p$ Hermitian positive definite matrix $\boldsymbol{W}$ has $\mathbb{C}W_p(\boldsymbol{I}_p, n)$, where $n \geq p$ are integers. Let $\boldsymbol{T} = (t_{ij})_{\substack{i=1,2,\ldots,p \\ j=1,2,\ldots,i}}$ be a $p \times p$ lower triangular complex matrix with real and positive diagonal elements and $\boldsymbol{Q} = (q_{ij})_{\substack{i=1,2,\ldots,p \\ j=i,i+1,\ldots,p}}$ a $p \times p$ upper triangular complex matrix with real and positive diagonal elements such that $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{Q}\boldsymbol{Q}^*$. Then the elements of $\boldsymbol{T}$ are all independent, $2t_{ii}^2$ has $\chi^2_{2(n+1-i)}$ $(i = 1, 2, \ldots, p)$, the central χ^2 distribution with the degree of freedom $n+1-i$, and t_{ij} has $\mathbb{C}N(0, 1)$ $(1 \leq j < i \leq p)$. Furthermore the elements of $\boldsymbol{Q}$ are all independent, $2q_{ii}^2$ has $\chi^2_{2(n-p+i)}$ $(i = 1, 2, \ldots, p)$ and q_{ij} has $\mathbb{C}N(0, 1)$ $(1 \leq i < j \leq p)$.*

Proof. From Mathai (1997, Theorem 3.7), the Jacobian of the transformation from $\boldsymbol{W}$ to $\boldsymbol{T}$ is given as

$$(d\boldsymbol{W}) = 2^p \left\{ \prod_{i=1}^{p} t_{ii}^{2(p-i)+1} \right\} \bigwedge_{1 \leq j \leq i \leq p} dt_{ij}.$$

Substituting this express in (1.1), we can find that the joint density function of the t_{ij} $(1 \leq j \leq i \leq p)$ is

$$\prod_{i=1}^{p} \left\{ \frac{\{2t_{ii}^2\}^{(n+1-i)-1} \exp\{-t_{ii}^2\}}{\Gamma(n+1-i)2^{n+1-i}} d(2t_{ii}^2) \prod_{j=1}^{i-1} \frac{1}{\pi} \exp\{-|t_{ij}|^2\} dt_{ij} \right\},$$

which is the product of the marginal density functions of $\chi^2_{2(n+1-i)}$ $(i = 1, 2, \ldots, p)$ and $\mathbb{C}N(0, 1)$ $(1 \leq j < i \leq p)$. The rest of the proof can be obtained from a manner similar to that described above. □

Theorem 2.1. (a). *The best estimator among* (2.1) *is given by the diagonal matrix* $\boldsymbol{D}_T = \mathbf{Diag}(d_1^T, d_2^T, \ldots, d_p^T)$, *where the* d_i^T $(i = 1, 2, \ldots, p)$ *is given as the solutions of the equation*

$$\boldsymbol{B}_T(d_1^T, d_2^T, \ldots, d_p^T)' = (n+p-1, n+p-3, \ldots, n-p+1)'.$$

Here $\boldsymbol{B}_T = (b_{ij}^T)_{\substack{i=1,2,\ldots,p \\ j=1,2,\ldots,p}}$ *is a* $p \times p$ *symmetric matrix with*

$$b_{ii}^T = (n+p+1-2i)(n+p+2-2i); \qquad b_{ij}^T = n+p+1-2j; \qquad \text{for } i < j.$$

(b). *The best estimator among* (2.2) *is given by the diagonal matrix* $\boldsymbol{D}_Q = \mathbf{Diag}(d_1^Q, d_2^Q, \ldots, d_p^Q)$, *where the* d_i^Q $(i = 1, 2, \ldots, p)$ *is given as the solutions of the equation*

$$\boldsymbol{B}_Q(d_1^Q, d_2^Q, \ldots, d_p^Q)' = (n-p+1, n-p+3, \ldots, n+p-1)'.$$

Here $\boldsymbol{B}_Q = (b_{ij}^Q)_{\substack{i=1,2,\ldots,p\\ j=1,2,\ldots,p}}$ *is a* $p \times p$ *symmetric matrix with*

$$b_{ii}^Q = (n-p+1+2i)(n-p+2+2i); \qquad b_{ij}^Q = n-p+1+2i; \qquad \textit{for } i<j.$$

Proof. We only give the proof for (a) because a minor modification of the proof gives that for (b). Since the risk function for $\widehat{\boldsymbol{\Sigma}}_T = \boldsymbol{T}\boldsymbol{D}_T\boldsymbol{T}^*$ does not depend on $\boldsymbol{\Sigma}$, we need only to compare for $\boldsymbol{\Sigma} = \boldsymbol{I}_p$:

$$\begin{aligned}\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}_T - \boldsymbol{I}_p)^2] &= \mathbb{E}[\mathrm{Tr}\,(\boldsymbol{T}^*\boldsymbol{T}\boldsymbol{D}_T\boldsymbol{T}^*\boldsymbol{T}\boldsymbol{D}_T)] - 2\mathbb{E}[\mathrm{Tr}\,(\boldsymbol{T}^*\boldsymbol{T}\boldsymbol{D}_T)] + p \\ &=: A_1 - 2A_2 + p,\end{aligned}$$

where the expectation is taken with respect to $\mathbb{C}W_p(\boldsymbol{I}_p,\, n)$. Using Bartlett's decomposition and the fact that the forth central moment of the absolute value of a standard complex normal random variable is two, we can see that

$$\begin{aligned} A_1 &= \mathbb{E}\Bigg[\sum_{i=1}^p \Bigg\{\Big\{|t_{ii}|^4 + 2\sum_{k=i+1}^p |t_{ii}|^2\cdot|t_{ki}|^2 + \sum_{k=i+1}^p |t_{ki}|^4 \\ &\qquad + 2\sum_{k=i+1}^p\sum_{l=k+1}^p |t_{ki}|^2|t_{li}|^2\Big\}(d_i^T)^2 + 2\sum_{j=1}^{i-1}\sum_{k=i}^p |t_{ki}|^2\cdot|t_{kj}|^2 d_i^T d_j^T\Bigg\}\Bigg] \\ &= \mathbb{E}\Bigg[\sum_{i=1}^p \Bigg\{(n+p+1-2i)(n+p+2-2i)(d_i^T)^2 \\ &\qquad + 2\sum_{j=1}^{i-1}(n+p+1-2i)d_i^T d_j^T\Bigg\}\Bigg]. \end{aligned}$$

We perform a calculation similar to that described above to get

$$A_2 = \sum_{i=1}^{p}\sum_{k=i}^{p}\mathbb{E}[|t_{ki}|^2]d_i^T = \sum_{i=1}^{n}(n+p+1-2i)d_i^T.$$

From these expressions we can find that

$$\begin{aligned}\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}_T - \boldsymbol{I}_p)^2] \quad &= \quad \sum_{i=1}^{p}\Big\{(n+p+1-2i)(n+p+2-2i)(d_i^T)^2 \\ &+2\sum_{j=1}^{i-1}(n+p+1-2i)d_i^T d_j^T - 2(n+p+1-2i)d_i^T\Big\}.\end{aligned}$$

Differentiating the right hand side of the above equation with respect to the d_i^T $(i = 1,\, 2,\, \ldots,\, p)$ and equating to zero, we can see that the minimizers are given by the solutions of the equation. □

Remark 2.1. Consider estimators which satisfy

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{P}\boldsymbol{W}\boldsymbol{P}^*) = \boldsymbol{P}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{P}^*$$

for any $p \times p$ lower(upper) triangular matrix $\boldsymbol{P}$. Then these estimators have the form

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{T}\boldsymbol{V}\boldsymbol{T}^* \qquad (\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{Q}\boldsymbol{V}\boldsymbol{Q}^*) \tag{2.3}$$

where $\boldsymbol{T}\,(\boldsymbol{Q})$ is a $p \times p$ lower(upper) triangular matrix such that $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}^*\,(\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{Q}^*)$ and $\boldsymbol{V}$ is a $p \times p$ Hermitian constant matrix. By an argument similar to that in the proof of Theorem 2.1, the estimator given by Theorem 2.1 is still best among the estimators of the form (2.3). Since the group of upper triangular matrices is solvable, the Hunt-Stein Theorem implies that the estimator given in Theorem 2.1 is minimax under the loss function L_1.

3. Unbiased risk estimate for unitary invariant estimators and alternative estimators

Write $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}^*$, where $\boldsymbol{U}$ is a $p \times p$ unitary matrix such that $\boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{I}_p$, and $\boldsymbol{L} = \mathbf{Diag}(\ell_1, \ell_2, \ldots, \ell_p)$ with $\ell_1 > \ell_2 > \cdots > \ell_p$ being the ordered eigenvalues of $\boldsymbol{W}$. Note that the set of $\boldsymbol{W}$ with two or more equal eigenvalues is a set of zero Lebesgue measure[see Farrell (1985, page 74)]. Write $\mathbb{R}^p_> = \{(a_1, a_2, \ldots, a_p)' \in \mathbb{R}^p : a_1 > a_2 > \cdots > a_p > 0\}$. We consider a class of estimators of the form

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{U}\mathbf{Diag}(\varphi_1, \varphi_2, \ldots, \varphi_p)\boldsymbol{U}^*, \tag{3.1}$$

where the $\varphi_k := \varphi_k(\boldsymbol{L})$ $(k = 1, 2, \ldots, p)$ is a differentiable real-valued function on $\mathbb{R}^p_>$. The estimators (3.1) are complex analogues of those for the multivariate real normal covariance matrix discussed in Dey and Srinivasan (1985), Stein (1977), and Haff (1988, 1991). Note that the risk for the estimators (3.1) generally depends on $\boldsymbol{\Sigma}$. To compare the estimators (3.1) with scalar multiple of the empirical covariance matrix which has a constant risk, we employ the unbiased risk estimate approach due to Stein (1977). To derive the unbiased risk estimate for the estimators (3.1), we need integration by parts formula and calculus on eigenstructures for the nonsingular complex Wishart matrices.

To describe integration by parts for $\boldsymbol{W}$, we set the following notation: Let $\boldsymbol{G} := \boldsymbol{G}(\boldsymbol{W}) = (g_{ij})_{\substack{i=1,2,\ldots,p \\ j=1,2,\ldots,p}}$ be a $p \times p$ matrix of complex entries. The (i, j)-element g_{ij} is a function of $\boldsymbol{W} = (w_{ij})_{\substack{i=1,2,\ldots,p \\ j=1,2\ldots,p}}$. For the $p \times p$ Hermitian matrix $\boldsymbol{W}$, let $\boldsymbol{D}_W = (d_W^{jk})$ be a $p \times p$ operator matrix, the (j, k)-element of which is given by

$$d_W^{jk} = \frac{1}{2}(1+\delta_{jk})\left\{\frac{\partial}{\partial(\mathrm{Re}\, w_{jk})} + \sqrt{-1}(1-\delta_{jk})\frac{\partial}{\partial(\mathrm{Im}\, w_{jk})}\right\}, \tag{3.2}$$

for $j,\ k = 1,\ 2,\ \ldots,\ p$. Here, δ_{jk} is the Kronecker delta ($=1$ if $j=k$ and $=0$ if $j \neq k$). Thus the $(j,\ k)$-element of $\boldsymbol{D}_W\boldsymbol{G}$ is

$$\sum_{l=1}^{p} d_W^{jl} g_{lk} = \frac{1}{2}(1+\delta_{jl})\sum_{l=1}^{p}\left\{\frac{\partial g_{lk}}{\partial(\mathrm{Re}\, s_{jl})} + \sqrt{-1}(1-\delta_{jl})\frac{\partial g_{lk}}{\partial(\mathrm{Im}\, s_{jl})}\right\}.$$

Next two lemmas were given by Svensson (2004) and Svensson and Lundberg (2004). Note that the operator (3.2) is a slightly different version from that in Svensson (2004) so that expressions in the next two lemmas are changed correspondingly change[see Konno (2007b)].

Lemma 3.1. *Assume that each entry of* $\boldsymbol{G}$ *is a partially differentiable function with respect to* $\mathrm{Re}\, w_{jk}$ *and* $\mathrm{Im}\, w_{jk}$*,* $j,\ k = 1,\ 2,\ \ldots,\ p$*. Under the regularity conditions stated in* Konno (2007b)*, we have*

$$\mathbb{E}[\mathrm{Tr}(\boldsymbol{G}\boldsymbol{\Sigma}^{-1})] = \mathbb{E}[(n-p)\mathrm{Tr}(\boldsymbol{G}\boldsymbol{W}^{-1}) + \mathrm{Tr}(\boldsymbol{D}_W\boldsymbol{G})].$$

Lemma 3.2. *Decompose* $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}^*$ *in which* $\boldsymbol{U} = (u_{jk})_{\substack{j=1,2,\ldots,p\\k=1,2,\ldots,p}}$ *is a* $p \times p$ *unitary matrix and* $\boldsymbol{W} = \mathbf{Diag}(\ell_1,\ \ell_2,\ \ldots,\ \ell_p)$ *with* $\ell_1 > \ell_2 > \cdots > \ell_p > 0$.

Then we have, for $j,\ k,\ s,\ t = 1,\ 2,\ \ldots,\ p$,

$$\begin{aligned} d_W^{st}\ell_k &= u_{sk}\bar{u}_{tk}, \\ d_W^{st}u_{jk} &= \sum_{a\neq k}\frac{u_{ja}\bar{u}_{ta}u_{sk}}{\ell_k-\ell_a}, \\ d_W^{st}\bar{u}_{jk} &= \sum_{a\neq k}\frac{\bar{u}_{ja}u_{sa}\bar{u}_{tk}}{\ell_k-\ell_a}, \end{aligned}$$

in which d_W^{st} is given by (3.2).

Lemma 3.3. *Decompose $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}^*$ in which $\boldsymbol{U} = (u_{jk})_{\substack{j=1,2,\ldots,p\\k=1,2,\ldots,p}}$ is a $p\times p$ unitary matrix and $\boldsymbol{W} = \mathbf{Diag}(\ell_1,\ \ell_2,\ \ldots,\ \ell_p)$ with $\ell_1 > \ell_2 > \cdots > \ell_p > 0$. Let $\boldsymbol{\Phi} := \boldsymbol{\Phi}(\boldsymbol{L}) = \mathbf{Diag}(\varphi_1(\boldsymbol{L}),\ \varphi_2(\boldsymbol{L}),\ \ldots,\ \varphi_p(\boldsymbol{L}))$, where the $\varphi_k(\boldsymbol{L}) =: \varphi_k$ $(k = 1,\ 2\ \ldots,\ p)$ is a differentiable real-valued function on $\mathbb{R}^p_>$ with $\varphi_k(\mathbf{L}) \geq 0$. Under the regularity conditions, we have*

$$\mathbb{E}[\mathrm{Tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{U}^*)] = \mathbb{E}\Big[\sum_{k=1}^{p}\Big\{(n-p)\frac{\varphi_k}{\ell_k} + \frac{\partial\varphi_k}{\partial\ell_k} + \sum_{b\neq k}^{p}\frac{\varphi_k-\varphi_b}{\ell_k-\ell_b}\Big\}\Big].$$

Next lemma is useful for the derivation of the unbiased risk estimate for unitarily invariant estimators under the quadratic loss function L_1.

Lemma 3.4. *Let $\varphi_1(\boldsymbol{L}),\ \varphi_2(\boldsymbol{L}),\ \ldots,\ \varphi_p(\boldsymbol{L})$ be differentiable real-valued functions on $\mathbb{R}^p_>$ with $\varphi_k := \varphi_k(\boldsymbol{L}) \geq 0\,(k = 1,\ 2,\ \ldots,\ p)$ and let $\boldsymbol{\Phi} := \boldsymbol{\Phi}(\boldsymbol{L}) = \mathbf{Diag}(\varphi_1(\boldsymbol{L}),\ \varphi_2(\boldsymbol{L}),\ \ldots,\ \varphi_p(\boldsymbol{L}))$. Under suitable conditions, we have*

$$\mathbb{E}[\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}})] \;=\; \mathbb{E}[\mathrm{Tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}\boldsymbol{\Phi}^{(1)}\boldsymbol{U}^*)],$$

where $\widehat{\mathbf{\Sigma}} = \boldsymbol{U}\mathbf{\Phi}\boldsymbol{U}^*$ *and* $\mathbf{\Phi}^{(1)} = \mathbf{Diag}(\varphi_1^{(1)}, \varphi_2^{(1)}, \ldots, \varphi_p^{(1)})$ *with*

$$\varphi_k^{(1)} = (n-p)\frac{\varphi_k^2}{\ell_k} + 2\varphi_k \frac{\partial \varphi_k}{\partial \ell_k} + 2\varphi_k \sum_{b \neq k}^{p} \frac{\varphi_k - \varphi_b}{\ell_k - \ell_b} \text{ for } k = 1, 2, \ldots, p. \quad (3.3)$$

Proof. Using Lemma 3.1, we can find that

$$\mathbb{E}[\mathrm{Tr}(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}})] = \mathbb{E}[(n-p)\mathrm{Tr}\,(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}\boldsymbol{W}^{-1}\widehat{\mathbf{\Sigma}}) + \mathrm{Tr}\,(\boldsymbol{D}_W\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}})].$$

We evaluate the expectation of $\mathrm{Tr}\,(\boldsymbol{D}_W\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}})$. To this end, write $\widehat{\mathbf{\Sigma}} = (\widehat{\sigma}_{ij})_{\substack{i=1,2\ldots,p \\ j=1,2,\ldots,p}}$ and $\mathbf{\Sigma}^{-1} = (\sigma^{ij})_{\substack{i=1,2\ldots,p \\ j=1,2,\ldots,p}}$. Then we have

$$\begin{aligned} \mathrm{Tr}\,(\boldsymbol{D}_W\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}) &= \sum_{i,c_1,c_2,c_3=1}^{p} d_W^{ic_1}\widehat{\sigma}_{c_1c_2}\sigma^{c_2c_3}\widehat{\sigma}_{c_3i} \\ &= \sum_{c_2,c_3=1}^{p} \Big\{ \sigma^{c_2c_3}T_{c_3c_2}^{(1)} + \sigma^{c_2c_3}T_{c_3c_2}^{(2)} \Big\}, \end{aligned} \quad (3.4)$$

where $T_{c_3c_2}^{(1)} = \sum_{i,c_1=1}^{p}\widehat{\sigma}_{c_1c_2}(d_W^{ic_1}\widehat{\sigma}_{c_3i})$ and $T_{c_3c_2}^{(2)} = \sum_{i,c_1=1}^{p}\widehat{\sigma}_{c_3i}(d_W^{ic_1}\widehat{\sigma}_{c_1c_2})$. Using

Lemma 3.2, we can find that

$$
\begin{aligned}
T^{(1)}_{c_3c_2} &= \sum_{i,\,c_1=1}^{p} \widehat{\sigma}_{c_1c_2} d_W^{ic_1} \{\sum_{c_4=1}^{p} u_{c_3c_4}\varphi_{c_4}\bar{u}_{ic_4}\} \\
&= \sum_{i,\,c_1,\,c_4=1}^{p} \Big\{ \widehat{\sigma}_{c_1c_2}\varphi_{c_4}\bar{u}_{ic_4} \sum_{b\neq c_4}^{p} \frac{u_{c_3b}\bar{u}_{c_1b}u_{ic_4}}{\ell_{c_4}-\ell_b} + \widehat{\sigma}_{c_1c_2}\varphi_{c_4}u_{c_3c_4} \sum_{b\neq c_4}^{p} \frac{\bar{u}_{ib}u_{ib}\bar{u}_{c_1c_4}}{\ell_{c_4}-\ell_b} \\
&\quad + \sum_{m=1}^{p} \widehat{\sigma}_{c_1c_2}u_{c_3c_4}\bar{u}_{ic_4}u_{im}\bar{u}_{c_1m}\frac{\partial\varphi_{c_4}}{\partial\ell_m} \Big\} \\
&= \{\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_1-\varphi_b}{\ell_1-\ell_b}+\frac{\partial\varphi_1}{\partial\ell_1},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_p-\varphi_b}{\ell_p-\ell_b}+\frac{\partial\varphi_p}{\partial\ell_p})\boldsymbol{U}^*\widehat{\boldsymbol{\Sigma}}\}_{c_3c_2}.
\end{aligned}
$$

Here we denote by $\{\boldsymbol{A}\boldsymbol{B}\}_{ij}$ the $(i,\,j)$-element of the products of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. We perform a calculation similar to that described above to get

$$
T^{(2)}_{c_3c_2} = \{\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_1-\varphi_b}{\ell_1-\ell_b}+\frac{\partial\varphi_1}{\partial\ell_1},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_p-\varphi_b}{\ell_p-\ell_b}+\frac{\partial\varphi_p}{\partial\ell_p})\boldsymbol{U}^*\}_{c_3c_2}.
$$

Putting these two expressions in (3.4), we can find the desired result. □

Theorem 3.1. *For* $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{U}\mathbf{Diag}(\varphi_1,\,\varphi_2,\,\ldots,\,\varphi_p)\boldsymbol{U}^*$, *we have*

$$
\begin{aligned}
\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2] &= \mathbb{E}\Big[\sum_{k=1}^{p}\Big\{(n-p)(\frac{\varphi_k^{(1)}}{\ell_k}-2\frac{\varphi_k}{\ell_k})+(\frac{\partial\varphi_k^{(1)}}{\partial\ell_k}-2\frac{\partial\varphi_k}{\partial\ell_k}) \\
&\quad + \sum_{b\neq k}^{p}\frac{(\varphi_k^{(1)}-2\varphi_k)-(\varphi_b^{(1)}-2\varphi_b)}{\ell_k-\ell_b}+1\Big\}\Big],
\end{aligned}
$$

where the $\varphi_k^{(1)}$ $(k=1,\,2,\,\ldots,\,p)$ *is given by* (3.3).

Proof. This theorem can be seen from Lemmas 3.3 and 3.4. □

Write $R(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathbb{E}[L_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$, the risk function for an estimator $\widehat{\boldsymbol{\Sigma}}$. Using Theorem 3.1, we evaluate risk for a scalar multiple of the empirical covariance matrix and alternative estimators which are complex analogues of estimators for the multivariate real normal covariance matrices given by Haff (1980).

Proposition 3.1. *Consider the form of estimators $\widehat{\boldsymbol{\Sigma}}_a = a\boldsymbol{W}$, where a is a positive constant. Then the best constant is given by $a = 1/(n+p)$ under the L_1 loss.*

Proof. Note that we have $\varphi_k^{(1)} = (n+p)a^2\ell_k$ for $\varphi_k = a\ell_k$ $(k = 1, 2, \dots, n)$. Hence, from Theorem 3.1, we can find that

$$
\begin{aligned}
R(\widehat{\boldsymbol{\Sigma}}_a, \boldsymbol{\Sigma}) &= n\{(n+p)a^2 - 2a\} + p \\
&= n(n+p)\bigl(a - \frac{1}{n+p}\bigr)^2 + \frac{n(p-1)+p^2}{n+p},
\end{aligned}
$$

which completes the proof. □

Proposition 3.2. *Put $a = (n+p)^{-1}$ and consider estimators of the form*

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}} &= \frac{1}{n+p}\boldsymbol{U}\mathbf{Diag}(\ell_1 + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}}, \ell_2 + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}}, \dots, \ell_p + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}})\boldsymbol{U}^* \\
&= \frac{1}{n+p}\bigl(\boldsymbol{W} + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}}\boldsymbol{I}_p\bigr)
\end{aligned}
$$

for a positive constant t. Then the estimators $\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}}$ improve upon $\widehat{\boldsymbol{\Sigma}}_a$ if $0 < t < 2(p-1)(n-p)/\{(n-p+1)(n-p+2)\}$ under the L_1 loss.

Proof. Apply Theorem 3.1 with $\varphi_k = a(\ell_k + t/\mathrm{Tr}\,\boldsymbol{W}^{-1})$ for $k = 1, 2, \dots, p$.

Then we can find that the $\varphi_k^{(1)}$ for this choice is given as

$$a\ell_k + 2a^2\Big(\frac{n}{\mathrm{Tr}\,\boldsymbol{W}^{-1}} + \frac{1}{\ell_k(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\Big)t + a^2\Big(\frac{n-p}{\ell_k(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{2}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\Big)t^2.$$

Therefore, noting that

$$\sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_k^{-1}-\ell_b^{-1}}{\ell_k-\ell_b} < 0; \qquad \text{and} \qquad \sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_k^{-2}-\ell_b^{-2}}{\ell_k-\ell_b} < 0,$$

we have

$$\begin{aligned} &R(\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}},\,\boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}_a,\,\boldsymbol{\Sigma}) \\ &\quad< a\mathbb{E}\Big[(n-p)\big\{2a\big(n+\frac{2\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t + a\frac{(n-p+2)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t^2 - 2t\big\} \\ &\qquad + \big\{2a\frac{(n+1)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t + a\frac{(n-p+2)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t^2 - \frac{2\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t\big\}\Big]. \end{aligned}$$

But the coefficients of $\{\mathrm{Tr}\,\boldsymbol{W}^{-2}/(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2\}t$ is evaluated as

$$\{2a(n-p)+2a(n+1)-2\}\frac{\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t < 2a(n-p)t,$$

from which it follows that

$$R(\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}},\,\boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}_a,\,\boldsymbol{\Sigma}) < a^2\big\{(n-p+1)(n-p+2)t^2 - 2(p-1)(n-p)t\big\}.$$

This completes the proof. □

Acknowledgements
This work was in part supported by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (C) (No.17500185).

References

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

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

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* 13:1581–1591.

Eaton, M. L. (1989). *Group Invariance Applications in Statistics.* NSF-CBMS Regional Conference Series in Probability and Statistics 1. Hayward, CA: Inst. Math. Statist.

Farrell, R. H. (1985). *Multivariate Calculation.* New York:Springer-Verlag.

Giri, N. C. (2004). *Multivariate Statistical Analysis*, 2nd ed. New York:Marcel Dekker.

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.* 34:152-178.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* 8:586-597.

Haff, L. R. (1988). The variational form of certain Bayes estimators. Unpublished manuscript.

Haff, L. R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* 19:1163-1190.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1:361-380. Berkeley:Univ. California Press.

Khatri, C. G. (1965). Statistical analysis based on a certain multivariate complex distribution. *Ann. Math. Statist.* 36:98-114.

Konno, Y. (1992). Improved estimation of matrix of normal mean and eigenvalues in the multivariate F-distribution. Doctoral dissertation. University of Tsukuba.

Konno, Y. (2007a). Estimation of normal covariance matrices parametrized by irreducible symmetric cones under Stein's loss. *J. Multivariate Anal.* 98:295-316.

Konno, Y. (2007b). Improving on the sample covariance matrix for a complex elliptically contoured distribution. *J. Statist. Plan. Infer.* 137:2475-2486.

Kubakawa, T. and Konno, Y. (1990). Estimating the covariance matrix and the generalized variance under a symmetric loss. *Ann. Inst. Statist. Math.* 42:331–343.

Kubokawa, T. and Srivastava, M. S. (1999). Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution. *Ann. Statist.* 27:600–609.

Loh, W. L. (1988). Estimating covariance matrices. Ph.D. thesis. Stanford University.

Loh, W. L. (1991). Estimating covariance matrices. *Ann. Statist.* 19:293-296.

Maiward, D. and Kraus, D. (1994). Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices. *Radar, Sonar and Navigation, IEEE Proceedings* 147:162-168.

Mathai, A. M. (1997). *Jacobian of Matrix Transformations and Functions of Matrix Argument.* Singapore:World Scientific.

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

Olkin, I. and Selliah, J. B. (1977). Estimating covariances in a multivariate normal distribution. *Statistical Decision Theory and Related Topics, II*(Proc. Sympos., Purdue Univ., Lafayette, Ind., 1976), 313–326. New York:Academic

Press.

Ratnaraja, T. and Vaillancourt, R. (2005). Complex singular Wishart matrices and applications. *Comput. Math. Appl.* 50:399-411.

Srivastava, M. S. (1965). On the complex Wishart distribution. *Ann. Math. Statist.* 36:313-315.

Stein, C. (1977). 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.), Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division 74, 4-65 (In Russia). English translation of this article is available in J. Sov. Math. 34 (1986) 1373-1403.

Svensson, L. (2004). A useful identity for complex wishart forms. Technical reports, Department of Signals and Systems, Chalmers University of Technology, December 2004.

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

Tague, J. A. and Caldwell, C. (1994). Expectations of useful complex Wishart forms. *Multidimens. Syst. Signal Process.* 5:263-279.

A Appendix: Detail calculation for Sections 2 and 3

This is a supplement to the manuscript for reviewers' convenience(only for review process).

A.1 Supplement to the proof of Lemma 2.1

Similarly, from Mathai (1997, Theorem 3.7), the Jacobian of the transformation from $\boldsymbol{W}$ to $\boldsymbol{Q}$ is given as

$$(d\boldsymbol{W}) = 2^p \left\{ \prod_{i=1}^{p} q_{ii}^{2(i-1)+1} \right\} \bigwedge_{1\leq i\leq j\leq p} dq_{ij}.$$

Substituting this express in (1.1), we can find that the joint density function of the q_{ij} $(1 \leq i \leq j \leq p)$ is

$$\prod_{i=1}^{p} \left\{ \frac{\{2q_{ii}^2\}^{(n-p+i)-1} \exp\{-q_{ii}^2\}}{\Gamma(n-p+i)2^{n-p+i}} d(2q_{ii}^2) \prod_{j=1}^{i-1} \frac{1}{\pi} \exp\{-|q_{ij}|^2\} dq_{ij} \right\},$$

which is the product of the marginal density functions of $\chi^2_{2(n-p+i)}$ $(i = 1,\, 2,\, \ldots,\, p)$ and $\mathbb{C}N(0,\, 1)$ $(1 \leq j < i \leq p)$. □

A.2 Calculation for Theorem 2.1

Calculation of A_1:

$$\begin{aligned}
A_1 &= \mathbb{E}\Big[\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=\max(i,j)}^{p}\sum_{l=\max(i,j)}^{p}\bar{t}_{ki}t_{kj}\bar{t}_{lj}t_{li}d_i^T d_j^T\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{p}\Big\{\big\{|t_{ii}|^4 + 2\sum_{k=i+1}^{p}|t_{ii}|^2\cdot|t_{ki}|^2 + \sum_{k=i+1}^{p}|t_{ki}|^4 \\
&\quad +2\sum_{k=i+1}^{p}\sum_{l=k+1}^{p}|t_{ki}|^2|t_{li}|^2\big\}(d_i^T)^2 + 2\sum_{j=1}^{i-1}\sum_{k=i}^{p}|t_{ki}|^2\cdot|t_{kj}|^2 d_i^T d_j^T\Big\}\Big] \\
&= \sum_{i=1}^{p}\Big\{\big\{(n+1-i)(n+2-i)+2(p-i)(n+1-i)+2(p-i) \\
&\quad +(p-i)(p-1-i)\big\}(d_i^T)^2 + 2\sum_{j=1}^{i-1}\{n+p+1-2i\}d_i^T d_j^T\Big\}\Big]\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{p}\Big\{(n+p+1-2i)(n+p+2-2i)(d_i^T)^2 \\
&\quad +2\sum_{j=1}^{i-1}(n+p+1-2i)d_i^T d_j^T\Big\}
\end{aligned}$$

Finally, note that

$$\begin{aligned}
&\frac{\partial}{\partial d_i^T}\sum_{k=1}^{p}\sum_{l=1}^{k-1}(n+p+1-2k)d_k^T d_l^T \\
&\qquad = \sum_{l=1}^{i-1}(n+p+1-2i)d_l^T + \sum_{k=i+1}^{p}(n+p+1-2k)d_k^T
\end{aligned}$$

to see that $b_{ij}^T\,(i<j)$ is given by $n+p+1-2j$. □

Proof of (b) in Theorem 2.1: Write

$$\begin{aligned}\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}_Q-\boldsymbol{I}_p)^2] &= \mathbb{E}[\mathrm{Tr}\,(\boldsymbol{Q}^*\boldsymbol{Q}\boldsymbol{D}_Q\boldsymbol{Q}^*\boldsymbol{Q}\boldsymbol{D}_Q)]-2\mathbb{E}[\mathrm{Tr}\,(\boldsymbol{Q}^*\boldsymbol{Q}\boldsymbol{D}_Q)]+p\\ &=: A_3-2A_4+p,\end{aligned}$$

where the expectation is taken with respect to $\mathbb{C}W_p(\boldsymbol{I}_p,\,n)$. Using Bartlett's decomposition and the fact that the forth central moment of standard complex normal distribution is two, we can see that that A_3 is given as

$$\begin{aligned}A_3 &= \mathbb{E}\Big[\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{k=\min(i,\,j)}\sum_{l=1}^{l=\min(i,\,j)}\bar{q}_{ki}q_{kj}\bar{q}_{lj}q_{li}d_i^Qd_j^Q\Big]\\ &= \mathbb{E}\Big[\sum_{i=1}^{p}\Big\{\{|q_{ii}|^4+2\sum_{k=1}^{i-1}|q_{ii}|^2\cdot|q_{ki}|^2+\sum_{k=1}^{i-1}|q_{ki}|^4\\ &\quad +2\sum_{k=1}^{i-1}\sum_{l=1}^{k-1}|q_{ki}|^2|q_{li}|^2\}(d_i^Q)^2+2\sum_{j=1}^{i-1}\sum_{k=1}^{j}|q_{ki}|^2\cdot|q_{kj}|^2d_i^Qd_j^Q\Big\}\Big]\\ &= \sum_{i=1}^{p}\Big\{\{(n-p+i)(n-p+1+i)+2(i-1)(n-p+i)+2(i-1)\\ &\quad +(i-1)(i-2)\}(d_i^Q)^2+2\sum_{j=1}^{i-1}\{n-p-1+2j\}d_i^Qd_j^Q\Big\}\Big]\Big]\\ &= \mathbb{E}\Big[\sum_{i=1}^{p}\Big\{(n-p-1+2i)(n-p+2i)(d_i^Q)^2\\ &\quad +2\sum_{j=1}^{i-1}(n-p-1+2j)d_i^Qd_j^Q\Big\}.\end{aligned}$$

Similarly we have

$$A_4=\sum_{i=1}^{p}\sum_{k=1}^{i}|q_{ki}|^2d_i^Q=\sum_{i=1}^{n}(n-p-1+2i)d_i^Q.$$

From these expressions we find that

$$\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}_Q - \boldsymbol{I}_p)^2] = \sum_{i=1}^{p}\Big\{(n-p-1+2i)(n-p+2i)(d_i^Q)^2 \\ +2\sum_{j=1}^{i-1}(n-p-1+2j)d_i^Q d_j^Q - 2(n-p-1+2i)d_i^Q\Big\}.$$

Differentiating the right hand side of the above equation with respect to the d_i^Q $(i = 1, 2, \ldots, p)$ and equating to zero, we can see that the minimizers are given by the solutions of the equation. Note that

$$\frac{\partial}{\partial d_i^T}\sum_{k=1}^{p}\sum_{l=1}^{k-1}(n+p+1-2l)d_k^T d_l^T \\ = \sum_{l=1}^{i-1}(n+p+1-2l)d_l^T + \sum_{k=i+1}^{p}(n+p+1-2i)d_k^T$$

to see that b_{ij}^T $(i < j)$ is given by $n + p + 1 - 2i$. □

A.3 Calculation for Lemma 3.3

Calculation of $T^{(1)}_{c_3c_2}$:

$$
\begin{aligned}
T^{(1)}_{c_3c_2} &= \sum_{i,\,c_1=1}^{p} \widehat{\sigma}_{c_1c_2} d_W^{ic_1}\{\sum_{c_4=1}^{p} u_{c_3c_4}\varphi_{c_4}\bar{u}_{ic_4}\} \\
&= \sum_{i,\,c_1,\,c_4=1}^{p} \Big\{ \widehat{\sigma}_{c_1c_2}\varphi_{c_4}\bar{u}_{ic_4}(d_W^{ic_1}u_{c_3c_4}) + \widehat{\sigma}_{c_1c_2}\varphi_{c_4}u_{c_3c_4}(d_W^{ic_1}\bar{u}_{ic_4}) \\
&\quad + \sum_{m=1}^{p} \widehat{\sigma}_{c_1c_2}u_{c_3c_4}\bar{u}_{ic_4}\frac{\partial\varphi_{c_4}}{\partial\ell_m}(d_W^{ic_1}\ell_m)\Big\} \\
&= \sum_{i,\,c_1,\,c_4=1}^{p} \Big\{ \widehat{\sigma}_{c_1c_2}\varphi_{c_4}\bar{u}_{ic_4}\sum_{b\neq c_4}^{p}\frac{u_{c_3b}\bar{u}_{c_1b}u_{ic_4}}{\ell_{c_4}-\ell_b} + \widehat{\sigma}_{c_1c_2}\varphi_{c_4}u_{c_3c_4}\sum_{b\neq c_4}^{p}\frac{\bar{u}_{ib}u_{ib}\bar{u}_{c_1c_4}}{\ell_{c_4}-\ell_b} \\
&\quad + \sum_{m=1}^{p} \widehat{\sigma}_{c_1c_2}u_{c_3c_4}\bar{u}_{ic_4}u_{im}\bar{u}_{c_1m}\frac{\partial\varphi_{c_4}}{\partial\ell_m}\Big\} \\
&= \{\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_b}{\ell_b-\ell_1},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_b}{\ell_b-\ell_p})\boldsymbol{U}^*\widehat{\boldsymbol{\Sigma}}\}_{c_3c_2} \\
&\quad +\{\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_1}{\ell_1-\ell_b},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_p}{\ell_p-\ell_b})\boldsymbol{U}^*\widehat{\boldsymbol{\Sigma}}\}_{c_3c_2} \\
&\quad +\{\boldsymbol{U}\mathbf{Diag}(\frac{\partial\varphi_1}{\partial\ell_1},\,\ldots,\,\frac{\partial\varphi_p}{\partial\ell_p})\boldsymbol{U}^*\widehat{\boldsymbol{\Sigma}}\}_{c_3c_2} \\
&= \{\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_1-\varphi_b}{\ell_1-\ell_b}+\frac{\partial\varphi_1}{\partial\ell_1},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_p-\varphi_b}{\ell_p-\ell_b}+\frac{\partial\varphi_p}{\partial\ell_p})\boldsymbol{U}^*\widehat{\boldsymbol{\Sigma}}\}_{c_3c_2}.
\end{aligned}
$$

□

Calculation of $T^{(2)}_{c_3c_2}$:

$$\begin{aligned}
T^{(2)}_{c_3c_2} &= \sum_{i,\,c_1=1}^{p} \widehat{\sigma}_{c_3i} d^{ic_1}_{W} \{\sum_{c_4=1}^{p} u_{c_1c_4}\varphi_{c_4}\bar{u}_{c_2c_4}\} \\
&= \sum_{i,\,c_1,\,c_4=1}^{p} \Big\{ \widehat{\sigma}_{c_3i}\varphi_{c_4}\bar{u}_{c_2c_4}(d^{ic_1}_{W} u_{c_1c_4}) + \widehat{\sigma}_{c_3i}\varphi_{c_4}u_{c_1c_4}(d^{ic_1}_{W}\bar{u}_{c_2c_4}) \\
&\quad + \sum_{m=1}^{p} \widehat{\sigma}_{c_3i}u_{c_1c_4}\bar{u}_{c_2c_4}\frac{\partial\varphi_{c_4}}{\partial\ell_m}(d^{ic_1}_{W}\ell_m)\Big\} \\
&= \sum_{i,\,c_1,\,c_4=1}^{p} \Big\{ \widehat{\sigma}_{c_3i}\varphi_{c_4}\bar{u}_{c_2c_4}\sum_{b\neq c_4}^{p}\frac{u_{c_1b}\bar{u}_{c_1b}u_{ic_4}}{\ell_{c_4}-\ell_b} + \widehat{\sigma}_{c_3i}\varphi_{c_4}u_{c_1c_4}\sum_{b\neq c_4}^{p}\frac{\bar{u}_{c_2b}u_{ib}\bar{u}_{c_1c_4}}{\ell_{c_4}-\ell_b} \\
&\quad + \sum_{m=1}^{p} \widehat{\sigma}_{c_3i}u_{c_1c_4}\bar{u}_{c_2c_4}u_{im}\bar{u}_{c_1m}\frac{\partial\varphi_{c_4}}{\partial\ell_m}\Big\} \\
&= \{\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_1}{\ell_1-\ell_b},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_p}{\ell_p-\ell_b})\boldsymbol{U}^*\}_{c_3c_2} \\
&\quad + \{\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_b}{\ell_b-\ell_1},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_b}{\ell_b-\ell_p})\boldsymbol{U}^*\}_{c_3c_2} \\
&\quad + \{\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\mathbf{Diag}(\frac{\partial\varphi_1}{\partial\ell_1},\,\ldots,\,\frac{\partial\varphi_p}{\partial\ell_p})\boldsymbol{U}^*\}_{c_3c_2} \\
&= \{\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\mathbf{Diag}(\sum_{b\neq 1}\frac{\varphi_1-\varphi_b}{\ell_1-\ell_b}+\frac{\partial\varphi_1}{\partial\ell_1},\,\ldots,\,\sum_{b\neq p}\frac{\varphi_p-\varphi_b}{\ell_p-\ell_b}+\frac{\partial\varphi_p}{\partial\ell_p})\boldsymbol{U}^*\}_{c_3c_2}.
\end{aligned}$$

□

A.4 Calculation for Proposition 3.2

Put $a = 1/(n+p)$ and apply Theorem 3.1 with $\varphi_k = a(\ell_k + t/\mathrm{Tr}\,\boldsymbol{W}^{-1})$ $(k = 1, 2, \ldots, p)$. Then we have, for $k = 1, 2, \ldots, p$,

$$\begin{aligned}
\varphi_k^{(1)} &= \frac{(n-p)a^2}{\ell_k}\big(\ell_k + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}}\big)^2 + 2a^2\big(\ell_k + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}}\big)\big(1 + \frac{t}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big) \\
&\quad +2a^2(p-1)\big(\ell_k + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{-1}}\big) \\
&= a\ell_k + 2a^2\big(\frac{n}{\mathrm{Tr}\,\boldsymbol{W}^{-1}} + \frac{1}{\ell_k(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t + a^2\big(\frac{n-p}{\ell_k(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{2}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\big)t^2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
& R(\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}}, \boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}_a, \boldsymbol{\Sigma}) \\
&= \mathbb{E}\Big[\sum_{k=1}^{n}\Big\{(n-p)\big\{2a^2\big(\frac{n}{\ell_k \mathrm{Tr}\,\boldsymbol{W}^{-1}} + \frac{1}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t \\
&\quad +a^2\big(\frac{n-p}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{2}{\ell_k^3(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\big)t^2 - \frac{2a}{\ell_k(\mathrm{Tr}\,\boldsymbol{W}^{-1})}t\big\} \\
&\quad +\big\{2a^2\big(\frac{n}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{2}{\ell_k^3(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3} - \frac{1}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t \\
&\quad +a^2\big(\frac{2(n-p)}{\ell_k^3(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3} - \frac{n-p}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{6}{\ell_k^4(\mathrm{Tr}\,\boldsymbol{W}^{-1})^4} - \frac{4}{\ell_k^3(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\big)t^2 \\
&\quad -\frac{2a}{\ell_k^2(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t\big\} + \frac{2a^2}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\sum_{b\neq k}^{p}\frac{\ell_k^{-1} - \ell_b^{-1}}{\ell_k - \ell_b}\cdot t \\
&\quad +\big(\frac{a^2(n-p)}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\sum_{b\neq k}^{p}\frac{\ell_k^{-1} - \ell_b^{-1}}{\ell_k - \ell_b} + \frac{2a^2}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\sum_{b\neq k}^{p}\frac{\ell_k^{-2} - \ell_b^{-2}}{\ell_k - \ell_b}\big)t^2\Big\}\Big].
\end{aligned}$$

But we have (the signs below are clearly negative)

$$\begin{aligned}\sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_k^{-1}-\ell_b^{-1}}{\ell_k-\ell_b} &= \sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_b-\ell_k}{\ell_K\ell_b(\ell_k-\ell_b)} = -\sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{1}{\ell_k\ell_b} \\ &= -\sum_{k=1}^{p}\frac{1}{\ell_k}\sum_{b=1}^{p}\frac{1}{\ell_b}+\sum_{k=1}^{p}\frac{1}{\ell_k^2} \\ &= -(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2+\mathrm{Tr}\,\boldsymbol{W}^{-2}<0,\end{aligned}$$

and

$$\begin{aligned}\sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_k^{-2}-\ell_b^{-2}}{\ell_k-\ell_b} &= \sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_b-\ell_k}{\ell_k^2\ell_b^2(\ell_k-\ell_b)} = -\sum_{k=1}^{p}\sum_{b\neq k}^{p}\frac{\ell_k+\ell_b}{\ell_k^2\ell_b^2} \\ &= -\sum_{k=1}^{p}\frac{1}{\ell_k}\sum_{b=1}^{p}\frac{1}{\ell_b^2}+\sum_{k=1}^{p}\frac{1}{\ell_k^3}-\sum_{k=1}^{p}\frac{1}{\ell_k^2}\sum_{b=1}^{p}\frac{1}{\ell_b}+\sum_{k=1}^{p}\frac{1}{\ell_k^3} \\ &= -2(\mathrm{Tr}\,\boldsymbol{W}^{-1})\mathrm{Tr}\,\boldsymbol{W}^{-2}+2\mathrm{Tr}\,\boldsymbol{W}^{-3}<0.\end{aligned}$$

Therefore we have

$$
\begin{aligned}
&R(\widehat{\mathbf{\Sigma}}_{\mathrm{HF}},\,\mathbf{\Sigma}) - R(\widehat{\mathbf{\Sigma}}_{a},\,\mathbf{\Sigma}) \\
&< \mathbb{E}\Big[(n-p)\big\{2a^2\big(n+\frac{\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t + a^2\big(\frac{(n-p)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{2\mathrm{Tr}\,\boldsymbol{W}^{-3}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\big)t^2 \\
&\quad -2at\big\} + \big\{2a^2\big(\frac{n\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{2\mathrm{Tr}\,\boldsymbol{W}^{-3}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3} - \frac{\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t \\
&\quad +a^2\big(\frac{2(n-p)\mathrm{Tr}\,\boldsymbol{W}^{-3}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3} - \frac{(n-p)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2} + \frac{6\mathrm{Tr}\,\boldsymbol{W}^{-4}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^4} - \frac{4\mathrm{Tr}\,\boldsymbol{W}^{-3}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^3}\big)t^2 \\
&\quad -\frac{2a\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t\big\}\Big] \\
&< a\mathbb{E}\Big[(n-p)\big\{2a\big(n+\frac{\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}\big)t + a\frac{(n-p+2)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t^2 - 2t\big\} \\
&\quad +\big\{2a\frac{(n+1)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t + a\frac{(n-p+2)\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t^2 - \frac{2\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t\big\}\Big].
\end{aligned}
$$

Since the coefficients of $\{\mathrm{Tr}\,\boldsymbol{W}^{-2}/(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2\}t$ is evaluated as

$$2a(n-p) + 2a(n+1) - 2 = 2a(n-p) - 2a(p-1) < 2a(n-p),$$

we have

$$\{2a(n-p) + 2a(p+1) - 2\}\frac{\mathrm{Tr}\,\boldsymbol{W}^{-2}}{(\mathrm{Tr}\,\boldsymbol{W}^{-1})^2}t < 2a(n-p)t,$$

from which it follows that

$$R(\widehat{\mathbf{\Sigma}}_{\mathrm{HF}},\,\mathbf{\Sigma}) - R(\widehat{\mathbf{\Sigma}}_{a},\,\mathbf{\Sigma}) < a^2\big\{(n-p+2)(n-p+1)t^2 - 2(p-1)(n-p)t\big\}.$$

This completes the proof. □