Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Multivariate Analysis 94 (2005) 271–299

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Distribution of eigenvalues and eigenvectors of Wishart matrix when the population eigenvalues are infinitely dispersed and its application to minimax estimation of covariance matrix

Akimichi Takemura[a,*], Yo Sheena[b,1]

[a]*Department of Mathematical Informatics, University of Tokyo, Graduate School of Information Science and Technology, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[b]*Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA*

Received 1 July 2003
Available online 7 July 2004

Abstract

We consider the asymptotic joint distribution of the eigenvalues and eigenvectors of Wishart matrix when the population eigenvalues become infinitely dispersed. We show that the normalized sample eigenvalues and the relevant elements of the sample eigenvectors are asymptotically all mutually independently distributed. The limiting distributions of the normalized sample eigenvalues are chi-squared distributions with varying degrees of freedom and the distribution of the relevant elements of the eigenvectors is the standard normal distribution. As an application of this result, we investigate tail minimaxity in the estimation of the population covariance matrix of Wishart distribution with respect to Stein's loss function and the quadratic loss function. Under mild regularity conditions, we show that the behavior of a broad class of tail minimax estimators is identical when the sample eigenvalues become infinitely dispersed.

AMS subject classification: 62H10; 62C20

Keywords: Asymptotic distribution; Covariance matrix; Minimax estimator; Quadratic loss; Singular parameter; Stein's loss; Tail minimaxity

* Corresponding author.
E-mail address: takemura@stat.t.u-tokyo.ac.jp (A. Takemura)

[1] On leave from Department of Economics, Shinshu University, Japan.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2004.05.003

1. Introduction

Let $\boldsymbol{W} = (w_{ij})$ be distributed according to Wishart distribution $\boldsymbol{W}_p(n, \boldsymbol{\Sigma})$, where p is the dimension, n is the degrees of freedom $(n \geqslant p)$ and $\boldsymbol{\Sigma}$ is the covariance matrix. We are interested in the joint distribution of the eigenvalues and the eigenvectors of $\boldsymbol{W}$. Because the exact distribution in terms of hypergeometric function of matrix arguments is cumbersome to handle, various types of asymptotic approximations have been investigated in literature.

The usual large sample theory $(n \to \infty)$ of the sample eigenvalues and eigenvectors was given by Anderson [1] and developed by many authors. Higher order asymptotic expansions for the case of distinct population roots were given by Sugiura [21] and Muirhead and Chikuse [14]. For a review of related works see Section 3 of Muirhead [12] and Section 10.3 of Siotani et al. [20]. Large sample theory under non-normality was studied by Waternaux [26], Davis [3] and Tyler [24,25].

The null case $\boldsymbol{\Sigma} = \boldsymbol{I_p}$ is of particular interest and there are two other types of asymptotics, different from the usual large sample asymptotics. One approach is the tube method (see [10] and the references therein), which gives asymptotic expansion of the tail probability of the largest root of Wishart matrix. Another approach is related to the field of random matrix theory and gives asymptotic distribution of the largest root for large dimension p (see [5,7] and the references therein).

In this paper, we consider yet another type of asymptotics, where the population eigenvalues become infinitely dispersed. Denote the spectral decompositions of $\boldsymbol{W}$ and $\boldsymbol{\Sigma}$ by

$$\boldsymbol{W} = \boldsymbol{GLG}', \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma\Lambda\Gamma}', \tag{1}$$

where $\boldsymbol{G}, \ \boldsymbol{\Gamma}$ are $p \times p$ orthogonal matrices and $\boldsymbol{L} = \mathrm{diag}(l_1, \dots, l_p), \ \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$ are diagonal matrices with the eigenvalues $l_1 \geqslant \cdots \geqslant l_p > 0, \ \lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$ of $\boldsymbol{W}$ and $\boldsymbol{\Sigma}$, respectively. We use the notations $\boldsymbol{l} = (l_1, \dots, l_p)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$ hereafter. We say that the population eigenvalues become *infinitely dispersed* when

$$\rho = \rho(\boldsymbol{\Sigma}) = \max\left(\frac{\lambda_2}{\lambda_1}, \frac{\lambda_3}{\lambda_2}, \dots, \frac{\lambda_p}{\lambda_{p-1}}\right) \to 0. \tag{2}$$

This limiting process includes many cases. For example, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$ may be of the following forms:

$$\begin{aligned}
&a^k(c, ac, a^2c, \dots, a^{p-1}c), \quad a \downarrow 0,\\
&(a^{j-1}c, a^{j-2}c, \dots, ac, c, bc, \dots, b^{p-j}c), \quad a \uparrow \infty, \ b \downarrow 0,
\end{aligned}$$

where $k \in R$ and $c > 0$ are fixed. In the parameter space of $\boldsymbol{\Sigma}$, i.e. the set of positive definite covariance matrices, the case of infinitely dispersed population eigenvalues corresponds to an extreme boundary of the parameter space. Geometrically, when ρ is extremely small, the concentration ellipsoid $\{\boldsymbol{x} \in R^p \mid \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = c\}$ is very flat; namely the projection of the ellipsoid onto a plane formed by any two eigenvectors of $\boldsymbol{\Sigma}$ yields an elongated ellipse.

We investigate the asymptotic distribution of the sample eigenvalues $(l_1, \dots, l_p)$ and the sample eigenvectors $\boldsymbol{G}$ under the limiting process (2). We will prove that after appropriate standardization, $l_1, \dots, l_p$ and relevant elements of $\boldsymbol{G}$ are asymptotically all mutually independently distributed. The limiting distributions of the sample eigenvalues are chi-squared distributions with varying degrees of freedom and the distribution of the relevant elements of $\boldsymbol{G}$ is the standard normal distribution.

The exact distribution of the eigenvectors and eigenvalues of $\boldsymbol{W}$ are complicated. Other than the null case $\boldsymbol{\Sigma} = \boldsymbol{I_p}$, it usually needs the expression by zonal polynomials, whose explicit forms are not known in a general case. If we consider the population parameter $\boldsymbol{\Sigma}$ in terms of $\lambda_{i+1}/\lambda_i,\ i = 1, \ldots, p-1$, the case where the population eigenvalues are infinitely dispersed corresponds to an end point

$$\frac{\lambda_{i+1}}{\lambda_i} = 0, \quad 1 \leqslant i \leqslant p-1,$$

while the null case

$$\frac{\lambda_{i+1}}{\lambda_i} = 1, \quad 1 \leqslant i \leqslant p-1,$$

is another end point. In this sense, the limiting case of infinitely dispersed population eigenvalues is quite contrasting to the null case. The above distributional results are interesting since they give clear answer to the distributions under the extreme situation.

Another motivation of the above result is the investigation of the tail minimaxity in the sense of Berger [2] in the estimation problem of $\boldsymbol{\Sigma}$ of Wishart distribution with respect to Stein's loss function and the quadratic loss function. For the case of the estimation of a location vector, Berger [2] gave sufficient conditions of tail minimaxity in a general multivariate location family. Under mild regularity conditions, we show that the behavior of a broad class of tail minimax estimators is identical when the sample eigenvalues become infinitely dispersed. This corresponds to a necessary condition of tail minimaxity in the estimation of $\boldsymbol{\Sigma}$.

The organization of this paper is as follows. In Section 2, we derive asymptotic distributions of $\boldsymbol{l}$ and $\boldsymbol{G}$. In Section 3, we prove tail minimaxity result for estimation of $\boldsymbol{\Sigma}$. All proofs of the lemmas are given in Appendix A and some additional technical results are given in Appendix B.

2. Asymptotic distributions of eigenvalues and eigenvectors

In this section, we derive the asymptotic distribution of the sample eigenvalues and eigenvectors when the population eigenvalues become infinitely dispersed. After preparing a lemma (Lemma 1), we state the consistency of the sample eigenvectors in Theorem 1 and derive the asymptotic distribution of normalized eigenvectors and eigenvalues in Theorem 2. At the end of the section, we show a technical lemma (Lemma 2). Proofs of the lemmas are given in Appendix A.

First we prove the following lemma concerning the tightness of the distribution of $l_i/\lambda_i,\ i = 1, \ldots, p$.

Lemma 1. *For any* $\varepsilon > 0$ *there exist* $C_1, C_2,\ 0 < C_1 < C_2$, *such that*

$$P\left(C_1 < \frac{l_i}{\lambda_i} < C_2,\ 1 \leqslant \forall i \leqslant p\right) > 1 - \varepsilon \quad \forall \boldsymbol{\Sigma}.$$

Note that C_1, C_2 above do not depend on $\mathbf{\Sigma}$. In the proof of this lemma in Appendix A, C_1 is taken to be sufficiently small and C_2 is taken to be sufficiently large to guarantee uniformity. As shown in Theorem 2, for small ρ, the above probability can be approximately evaluated by product of χ^2 probabilities.

From Lemma 1 we can easily show that the sample eigenvalues become infinitely dispersed in probability, when the population eigenvalues become infinitely dispersed. We omit the proof of the following corollary.

Corollary 1. *Let*

$$r = r(\boldsymbol{W}) = \max\left(\frac{l_2}{l_1}, \frac{l_3}{l_2}, \ldots, \frac{l_p}{l_{p-1}}\right). \tag{3}$$

Then as $\rho = \rho(\mathbf{\Sigma}) = \max(\lambda_2/\lambda_1, \ldots, \lambda_p/\lambda_{p-1}) \to 0$,

$$r \xrightarrow{p} 0$$

in the sense that $\forall \varepsilon > 0, \exists \delta > 0$,

$$\rho(\mathbf{\Sigma}) < \delta \;\;\Rightarrow\;\; P(r(\boldsymbol{W}) > \varepsilon) < \varepsilon.$$

From Lemma 1 we can prove the consistency of sample eigenvectors as the population eigenvalues become infinitely dispersed. Consider the spectral decomposition of $\boldsymbol{W}$ and $\mathbf{\Sigma}$ in (1). Here we are only considering the case where the population eigenvalues $\lambda_1, \ldots, \lambda_p$ are all distinct. Even in the case of the distinct population eigenvalues, the population eigenvectors are determined up to their signs. Various convenient rules (e.g. non-negativeness of the diagonal elements of $\boldsymbol{G}$ or $\mathbf{\Gamma}$) are used to determine the signs of the eigenvectors in the case of distinct roots. For the sample $\boldsymbol{W}$ these rules determine the signs of eigenvectors with probability 1, because the boundary set of these rules (e.g the set of $\boldsymbol{G}$ with 0 diagonal elements) is of measure 0. However, for the population covariance matrix $\mathbf{\Sigma}$, it is cumbersome to state the consistency result for $\mathbf{\Sigma}$ in the boundary set. Here we prefer to identify two eigenvectors of opposite signs $\boldsymbol{\gamma}$ and $-\boldsymbol{\gamma}$. Let $\mathcal{O}(p)$ denote the set of orthogonal matrices and let

$$\mathcal{O}(p)/\{-1, 1\}^p$$

denote the quotient set, where two orthogonal matrices $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are identified if there exists a diagonal matrix $\boldsymbol{D} = \mathrm{diag}(\pm 1, \ldots, \pm 1)$ such that $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_2\boldsymbol{D}$. Write two arbitrary orthogonal matrices $\mathbf{\Gamma}_i,\ i = 1, 2$, as $\mathbf{\Gamma}_1 = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)$, $\mathbf{\Gamma}_2 = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$. The elements of $\mathcal{O}(p)/\{-1, 1\}^p$ are sometimes called *frames* in the field of geometric probability (see [8, Chapter 6]). The quotient topology of $\mathcal{O}(p)/\{-1, 1\}^p$ induced from $\mathcal{O}(p)$ can be metrized, for example, by the squared distance

$$d^2(\mathbf{\Gamma}_1, \mathbf{\Gamma}_2) = \sum_{i=1}^{p} \mathrm{tr}(\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i' - \boldsymbol{\beta}_i\boldsymbol{\beta}_i')^2 = 2\sum_{i=1}^{p}(1 - (\boldsymbol{\gamma}_i'\boldsymbol{\beta}_i)^2). \tag{4}$$

We say that $\boldsymbol{\Gamma}_n$ converges to $\boldsymbol{\Gamma}$ in $\mathcal{O}(p)/\{-1,1\}^p$ if $d(\boldsymbol{\Gamma}_n, \boldsymbol{\Gamma})$ converges to 0 as $n \to \infty$. Note that d is orthogonally left invariant; i.e.,

$$d(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) = d(\boldsymbol{G}\boldsymbol{\Gamma}_1, \boldsymbol{G}\boldsymbol{\Gamma}_2), \quad \boldsymbol{G} \in \mathcal{O}(p).$$

Furthermore, noting that $\boldsymbol{\Gamma}'{}_1\boldsymbol{\Gamma}_2 = (\gamma_i'\boldsymbol{\beta}_j)$ is an orthogonal matrix, it is easily shown that

$$d(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) \to 0 \quad \Leftrightarrow \quad (\gamma_i'\boldsymbol{\beta}_j)^2 \to 0, \quad 1 \leqslant \forall j < \forall i \leqslant p. \tag{5}$$

Now we can state the consistency of $\boldsymbol{G}$ in $\mathcal{O}(p)/\{-1,1\}^p$.

Theorem 1. *Let $\boldsymbol{W} = \boldsymbol{G}\boldsymbol{L}\boldsymbol{G}'$, $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$ be the spectral decompositions of $\boldsymbol{W}$ and $\boldsymbol{\Sigma}$. Then as $\rho = \rho(\boldsymbol{\Sigma}) \to 0$*

$$\boldsymbol{G} \xrightarrow{p} \boldsymbol{\Gamma}$$

in $\mathcal{O}(p)/\{-1,1\}^p$, in the sense that $\forall \varepsilon > 0, \exists \delta > 0$,

$$\rho(\boldsymbol{\Sigma}) < \delta \quad \Rightarrow \quad P_{\boldsymbol{\Sigma}}(d(\boldsymbol{G}, \boldsymbol{\Gamma}) > \varepsilon) < \varepsilon.$$

Proof. Let

$$\widetilde{\boldsymbol{W}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Gamma}'\boldsymbol{W}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\widetilde{\boldsymbol{G}}\boldsymbol{L}\widetilde{\boldsymbol{G}}'\boldsymbol{\Lambda}^{-\frac{1}{2}} \sim \boldsymbol{W}_p(n, \boldsymbol{I_p}),$$

where $\widetilde{\boldsymbol{G}} = (\widetilde{g}_{ij}) = \boldsymbol{\Gamma}'\boldsymbol{G}$. By (5) it suffices to show that $\widetilde{g}_{ij}^2 \xrightarrow{p} 0$, $1 \leqslant \forall j < \forall i \leqslant p$. Suppose $j < i$. Note that

$$\widetilde{w}_{ii} = (\widetilde{g}_{i1}^2 l_1 + \cdots + \widetilde{g}_{ip}^2 l_p)\lambda_i^{-1}.$$

Therefore,

$$\widetilde{g}_{ij}^2 \leqslant \widetilde{w}_{ii}\frac{\lambda_i}{l_j} = \widetilde{w}_{ii}\frac{\lambda_j}{l_j}\frac{\lambda_i}{\lambda_j} \leqslant \widetilde{w}_{ii}\frac{\lambda_j}{l_j}\rho. \tag{6}$$

Since $\widetilde{w}_{ii}$ is independent of $\boldsymbol{\Sigma}$, for any $\varepsilon > 0$, there exists M such that

$$P(\widetilde{w}_{ii} < M) > 1 - \varepsilon \quad \forall \boldsymbol{\Sigma}. \tag{7}$$

Besides from the result of Lemma 1, for any $\varepsilon > 0$, there exists C such that

$$P\left(\left|\frac{\lambda_j}{l_j}\right| < C\right) > 1 - \varepsilon \quad \forall \boldsymbol{\Sigma}. \tag{8}$$

From (7) and (8), we can easily prove

$$\widetilde{w}_{ii}\frac{\lambda_j}{l_j}\rho \xrightarrow{p} 0 \quad \text{as } \rho \to 0.$$

From this fact and (6) we have

$$\widetilde{g}_{ij}^2 \xrightarrow{p} 0 \quad \text{as } \rho \to 0, \quad 1 \leqslant \forall j < \forall i \leqslant p. \qquad \square$$

The next theorem deals with the asymptotic distributions of standardized sample eigenvalues and sample eigenvectors. Here again we have to deal with the problem of indeterminacy of the signs of the eigenvectors. Let $\mathbf{\Sigma} = \mathbf{\Gamma\Lambda\Gamma}'$ have distinct roots. We assume that the signs of the columns of $\mathbf{\Gamma}$ are chosen in some way and fixed. Let

$$\widetilde{\boldsymbol{G}} = (\widetilde{g}_{ij}) = \mathbf{\Gamma}'\boldsymbol{G}.$$

Following Anderson [1], we assume that the signs of the columns of $\boldsymbol{G}$ is determined by requiring that the diagonal elements $\widetilde{g}_{ii}$ of $\widetilde{\boldsymbol{G}}$ are non-negative. Then it follows from Theorem 1 that $\widetilde{\boldsymbol{G}}$ converges to the identity matrix $\boldsymbol{I_p}$ in probability in the ordinary sense as the population eigenvalues become infinitely dispersed. The correct normalization for the sample eigenvalues and relevant elements of the eigenvectors are given by

$$f_i = \frac{l_i}{\lambda_i}, \quad 1 \leqslant i \leqslant p,$$

$$q_{ij} = \widetilde{g}_{ij} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}} = \widetilde{g}_{ij} f_j^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}, \quad 1 \leqslant j < i \leqslant p.$$

Note that the relevant elements of $\widetilde{\boldsymbol{G}}$ are the elements in the lower triangular part of $\widetilde{\boldsymbol{G}}$ (not including the diagonals). The asymptotic distribution of f_i and q_{ij} is given in the following theorem. Although this theorem is technically a corollary of the following Lemma 2, emphasizing the context, we state the theorem first.

Theorem 2. *As* $\rho = \max(\lambda_2/\lambda_1, \ldots, \lambda_p/\lambda_{p-1}) \to 0$,

$$f_i \xrightarrow{d} \chi^2_{n-i+1}, \quad 1 \leqslant i \leqslant p,$$

$$q_{ij} \xrightarrow{d} N(0,1), \quad 1 \leqslant j < i \leqslant p$$

and f_i $(1 \leqslant i \leqslant p)$, q_{ij} $(1 \leqslant j < i \leqslant p)$ *are asymptotically mutually independently distributed*.

Proof. Let A denote the event

$$A: \; l_i/\lambda_i \leqslant \beta_{ii}, \quad 1 \leqslant \forall i \leqslant p, \quad \tilde{g}_{ij} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}} \leqslant \beta_{ij}, \quad 1 \leqslant \forall j < \forall i \leqslant p,$$

with arbitrary real numbers β_{ij}, $1 \leqslant j \leqslant i \leqslant p$. Define $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) = I(A)$, where $I(\cdot)$ is the indicator function. Then from the definition of $x_{\mathbf{\Gamma}}$ in (10)

$$\begin{aligned} x_{\mathbf{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) &= I(f_i \leqslant \beta_{ii},\ 1 \leqslant \forall i \leqslant p,\ q_{ij} \leqslant \beta_{ij},\ 1 \leqslant \forall j < \forall i \leqslant p) \\ &= \bar{x}_{\mathbf{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}). \end{aligned}$$

Note that $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) = I(A)$ satisfies inequality (11) with $a = 0$, $b = 1$. Therefore by the Lemma 2, we have the result for any $\mathbf{\Gamma}$. □

Even for the case of $p = 2$, the explicit form of distribution functions of f_i, $i = 1, 2$, and q_{21} is complicated. We carried out simulation studies to check the accuracy of Theorem 2 for the case of $p = 2$ for degrees of freedom $n = 10, 20$ and 50. The replication size for the simulation is 10,000.

Table 1 presents the probabilities $P(f_i \leqslant \chi^2_{n-i+1}(0.05))$ and Table 2 presents the probabilities $P(f_i \leqslant \chi^2_{n-i+1}(0.95))$, where $\chi^2_m(\alpha)$ denotes the lower α percentage point of χ^2 distribution with m degrees of freedom. In Table 3 the probabilities $P(z_{0.05} < q_{21} \leqslant z_{0.95})$ are presented, where z_α is the lower α percentage point of the standard normal distribution.

Table 1
Lower 5% probability $P(f_i \leqslant \chi^2_{n-i+1}(0.05))$

ρ	$n=10$ f_1	$n=10$ f_2	$n=20$ f_1	$n=20$ f_2	$n=50$ f_1	$n=50$ f_2
0.1	0.04449	0.05089	0.04679	0.05143	0.04827	0.05124
0.01	0.05006	0.04963	0.05011	0.05012	0.04979	0.05025
0.001	0.04992	0.05187	0.04977	0.04983	0.05121	0.04915
0.0001	0.05138	0.05077	0.04879	0.04974	0.04945	0.0502

Table 2
Lower 95% probability $P(f_i \leqslant \chi^2_{n-i+1}(0.95))$

ρ	$n=10$ f_1	$n=10$ f_2	$n=20$ f_1	$n=20$ f_2	$n=50$ f_1	$n=50$ f_2
0.1	0.94862	0.95321	0.9486	0.95247	0.94882	0.95024
0.01	0.95047	0.95084	0.94879	0.9509	0.94958	0.95045
0.001	0.9497	0.95072	0.95057	0.9517	0.94902	0.94964
0.0001	0.94957	0.94912	0.94996	0.95073	0.94915	0.95086

Table 3
Central 90% probability $P(z_{0.05} \leqslant q_{21} \leqslant z_{0.95})$

ρ	$n=10$	$n=20$	$n=50$
0.1	0.85475	0.86003	0.85958
0.01	0.89568	0.89544	0.89764
0.001	0.89956	0.8998	0.89886
0.0001	0.90019	0.89962	0.89771

We see that for $\rho = 0.1$, the asymptotic distribution already well approximates the exact distribution.

Concerning Theorem 2, we again discuss the indeterminacy of the signs of the eigenvectors. In this theorem we choose the signs of sample eigenvectors by requiring nonnegativeness of the diagonal elements of $\widetilde{\boldsymbol{G}} = \boldsymbol{\Gamma}'\boldsymbol{G}$. This choice of the signs depends on the predetermined signs of the population eigenvectors, which are unknown in the setting of estimation. Although this choice is customary in standard large sample asymptotics, it might not be totally satisfactory. If we insist on choosing the signs of the sample eigenvectors independently of $\boldsymbol{\Gamma}$, there seem to be two ways of dealing with the indeterminacy of the signs. One way is to specify the sign of one element from each column of $\boldsymbol{G}$. This determines the signs with probability 1. Then Theorem 2 holds except for $\boldsymbol{\Sigma}$ in the corresponding boundary set. An alternative way is to choose the signs of the columns of $\boldsymbol{G}$ randomly, i.e., independently of the values of the elements of $\boldsymbol{G}$. Then it can be shown that Theorem 2 holds for all $\boldsymbol{\Sigma}$.

In Theorem 2 the asymptotic distribution of $\widetilde{\boldsymbol{G}}$ is described by its elements in the strictly lower triangular part. Note that we have the correct number of random variables because $\dim \mathcal{O}(p) = p(p-1)/2$. As discussed in Appendix B, in a neighborhood of $\boldsymbol{I_p}$, $\boldsymbol{G}$ is determined by its strictly lower triangular part,

$$(g_{21},\ g_{31}, \ldots, g_{p1},\ g_{32}, \ldots, g_{p,p-1}) = \boldsymbol{u} = (u_{ij})_{1 \leqslant j < i \leqslant p}. \tag{9}$$

All the other elements $g_{ij},\ 1\leqslant i\leqslant j\leqslant p$, are C^∞ functions of $\boldsymbol{u}$ on some open set U such that $\mathbf{0} = (0,\ldots,0) \in U \subset R^{p(p-1)/2}$. We write $\boldsymbol{G}(\boldsymbol{u}) = (g_{ij}(\boldsymbol{u}))$ with $g_{ij}(\boldsymbol{u}) = u_{ij},\ 1\leqslant j < i\leqslant p$. By Taylor expansion of $g_{ij}(\boldsymbol{u})$ $(1\leqslant i\leqslant j\leqslant p)$, we can study the asymptotic distributions of the upper part of $\boldsymbol{G}$. It turns out that the result cannot be simply expressed because it depends on the individual rates of the convergence of the ratios $\lambda_{i+1}/\lambda_i,\ 1\leqslant i\leqslant p-1$, to zero. This point is also discussed in Appendix B.

Finally, we present a technical lemma concerning the convergence of the expectation of a function of $\boldsymbol{W}$ and $\boldsymbol{\lambda}$, which will be used in the next section. Actually Theorem 2 is its corollary. Fix $\boldsymbol{\Gamma} \in \mathcal{O}(p)$. For any function $x(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})$ we define a compound function, $x_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})$, as

$$x_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda}) = x(\boldsymbol{\Gamma}\boldsymbol{G}(\boldsymbol{u}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})),\boldsymbol{l}(\boldsymbol{f},\boldsymbol{\lambda}),\boldsymbol{\lambda}), \tag{10}$$

where

$$\boldsymbol{f} = (f_1,\ldots,f_p),\quad \boldsymbol{q} = (q_{ij})_{1\leqslant j<i\leqslant p},\quad \boldsymbol{u}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda}) = (q_{ij} f_j^{-\frac12}\lambda_j^{-\frac12}\lambda_i^{\frac12})_{1\leqslant j<i\leqslant p},$$
$$\boldsymbol{l}(\boldsymbol{f},\boldsymbol{\lambda}) = (f_1\lambda_1,\ldots,f_p\lambda_p).$$

The domain of $x_{\boldsymbol{\Gamma}}$ as a function of $(\boldsymbol{f},\boldsymbol{q})$ is

$$\{(\boldsymbol{f},\boldsymbol{q}) \mid \boldsymbol{u}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda}) \in U,\ f_1\lambda_1 > \cdots > f_p\lambda_p\},$$

which expands to $R^p \times R^{p(p-1)/2}$ as $\rho \to 0$. $x_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})$ describes the local behavior of $x(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})$ around $\boldsymbol{\Gamma}$ with the coordinate $(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})$. We call $x_{\boldsymbol{\Gamma}}$ *the local expression of x around* $\boldsymbol{\Gamma}$ hereafter. Though it is not easy to calculate the explicit form of $\boldsymbol{G}(\boldsymbol{u})$, hence $x_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})$, we only need to know the limit of $x_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})$ in the following lemma.

Lemma 2. *Fix $\boldsymbol{\Gamma} \in \mathcal{O}(p)$ and let $\rho = \rho(\boldsymbol{\Lambda}) \to 0$ in $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$. Assume that $x_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q},\boldsymbol{\lambda})$ converges to a function $\bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q})$ a.e. in $(\boldsymbol{f},\boldsymbol{q})$ as $\rho(\boldsymbol{\Lambda}) \to 0$. Furthermore, assume that*

$$\exists a < \frac{1}{2},\ \exists b > 0,\quad |x(\boldsymbol{\Gamma}\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})| \leqslant b\ \exp\{\mathrm{tr}(a\boldsymbol{G}\boldsymbol{L}\boldsymbol{G}'\boldsymbol{\Lambda}^{-1})\}\quad a.e.\ in\ (\boldsymbol{G},\boldsymbol{l}). \tag{11}$$

Then as $\rho(\boldsymbol{\Lambda}) \to 0$,

$$E[x(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})] \to E[\bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{f},\boldsymbol{q})],$$

where the expectation on the right-hand side is taken with respect to the asymptotic distribution of $(\boldsymbol{f},\boldsymbol{q})$ given in Theorem 2.

3. Application to tail minimaxity in estimation of covariance matrix

The main motivation for the asymptotic theory in the previous section is the investigation of tail minimaxity in the estimation of the covariance matrix. We will state the mathematical definition of tail minimaxity later. For the case of the estimation of a location vector, Berger [2] gave sufficient conditions of tail minimaxity in a general multivariate location family. In this section, we derive some necessary conditions for an estimator $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ to be tail minimax with respect to Stein's (entropy) loss function as well as the quadratic loss function.

First assuming mild regularity condition (Assumption 1), we show that estimated eigenvectors are consistent in Lemma 3 and prove in Theorem 3 that they converge to the population eigenvectors when the sample eigenvalues become infinitely dispersed.

Second under additional condition (Assumption 2), we examine the asymptotic risk of an estimator in Theorem 4, and as its corollary we give a necessary condition for an estimator to be tail minimax (Corollary 2).

Now we briefly prepare some notations of the covariance estimation. For a survey of the estimation problem of $\boldsymbol{\Sigma}$ or $\boldsymbol{\Sigma}^{-1}$ in $\boldsymbol{W}_p(n, \boldsymbol{\Sigma})$, see [16]. Stein's loss function is one of the most frequently used loss functions for the estimation of $\boldsymbol{\Sigma}$ and it is given by

$$L_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p.$$

The first minimax estimator was given by James and Stein [6]. It is defined by

$$\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}} = \boldsymbol{T}\operatorname{diag}(\delta_1^{\mathrm{JS}}, \ldots, \delta_p^{\mathrm{JS}})\boldsymbol{T}', \tag{12}$$

where $\boldsymbol{T}$ is the lower triangular matrix with positive diagonal elements satisfying $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}'$ and

$$\delta_i^{\mathrm{JS}} = \frac{1}{n + p + 1 - 2i}, \quad 1 \leqslant i \leqslant p. \tag{13}$$

This type of estimators, $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{TDT}'$, $\boldsymbol{D} = \operatorname{diag}(\delta_1, \ldots, \delta_p)$, are called *triangularly equivariant*; i.e., for any lower triangular matrix $\boldsymbol{S}$ with positive diagonal elements, $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{SWS}') = \boldsymbol{S}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{S}'$. The estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}$ has the constant minimax risk, which is given by

$$\bar{R}_1 = -\sum_{i=1}^{p} \log \delta_i^{\mathrm{JS}} - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}]. \tag{14}$$

Another important loss function is the quadratic loss function given by

$$L_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I_p})^2.$$

Define a $p \times p$ matrix $\boldsymbol{A} = (a_{ij})$ and a $p \times 1$ vector $\boldsymbol{b} = (b_i)$ by

$$a_{ij} = \begin{cases} (n + p - 2i + 1)(n + p - 2i + 3) & \text{if } i = j, \\ (n + p - 2i + 1) & \text{if } i > j, \\ (n + p - 2j + 1) & \text{if } j > i, \end{cases} \tag{15}$$

$$b_i = n + p + 1 - 2i, \quad i = 1, \ldots, p \tag{16}$$

and let

$$\boldsymbol{\delta} = (\delta_1^{\mathrm{OS}}, \ldots, \delta_p^{\mathrm{OS}})' = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{17}$$

Then the minimax risk is given by

$$\bar{R}_2 = p - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} = p - \boldsymbol{\delta}'\boldsymbol{A}\boldsymbol{\delta} \tag{18}$$

[15,17] and the estimator

$$\widehat{\boldsymbol{\Sigma}}^{\mathrm{OS}} = \boldsymbol{T}\ \mathrm{diag}(\delta_1^{\mathrm{OS}}, \ldots, \delta_p^{\mathrm{OS}})\boldsymbol{T}'$$

has the constant minimax risk, where $\boldsymbol{T}$ is defined as in (12).

Here we introduce the definition of tail minimaxity. We call an estimator $\widehat{\boldsymbol{\Sigma}}$ *tail minimax* with respect to L_i, $i = 1, 2$, if there exists $\delta = \delta(\widehat{\boldsymbol{\Sigma}}) > 0$ such that

$$\rho(\boldsymbol{\Sigma}) < \delta \quad \Rightarrow \quad E[L_i(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})] \leqslant \bar{R}_i,$$

where $\bar{R}_i$ is the minimax risk for L_i and $\rho(\boldsymbol{\Sigma})$ is given in (2). Obviously a minimax estimator is tail minimax.

Let $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{G}, \boldsymbol{l})$ be an estimator of $\boldsymbol{\Sigma}$ and let

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{H}(\boldsymbol{W})\boldsymbol{D}(\boldsymbol{W})\boldsymbol{H}'(\boldsymbol{W})$$

be the spectral decomposition of $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})$, where

$$\boldsymbol{H}(\boldsymbol{W}) \in \mathcal{O}(p), \quad \boldsymbol{D}(\boldsymbol{W}) = \mathrm{diag}(d_1(\boldsymbol{W}), \ldots, d_p(\boldsymbol{W})).$$

In accordance with the definition of $\boldsymbol{G}$, we determine the sign of $\boldsymbol{H}(\boldsymbol{W})$ by

$$(\boldsymbol{\Gamma}'\boldsymbol{H})_{ii} \geqslant 0, \quad 1 \leqslant \forall i \leqslant p.$$

We also use the notation

$$c_i(\boldsymbol{W}) = \frac{d_i(\boldsymbol{W})}{l_i} = \frac{d_i(\boldsymbol{G}, \boldsymbol{l})}{l_i}, \quad i = 1, \ldots, p.$$

An estimator of the form

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{G}\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{G}', \quad \boldsymbol{\Psi}(\boldsymbol{L}) = \mathrm{diag}(\psi_1(\boldsymbol{l}), \ldots, \psi_p(\boldsymbol{l})) \tag{19}$$

is called *orthogonally equivariant*; i.e., $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{G}\boldsymbol{W}\boldsymbol{G}') = \boldsymbol{G}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{G}',\ \forall \boldsymbol{G} \in \mathcal{O}(p)$. For orthogonally equivariant estimators we have

$$\boldsymbol{H}(\boldsymbol{W}) = \boldsymbol{G},$$
$$c_i(\boldsymbol{W}) = c_i(\boldsymbol{l}) = \frac{\psi_i(\boldsymbol{l})}{l_i}, \quad 1 \leqslant i \leqslant p.$$

Here, we mention the orthogonally equivariant minimax estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ derived independently by Stein and by Dey and Srinivasan [4]. This estimator is defined by

$$\psi_i(\boldsymbol{l}) = l_i \delta_i^{\mathrm{JS}}, \quad 1 \leqslant i \leqslant p,$$

where δ_i^{JS} is given in (13). $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ is of simple form but has substantially better risk than the M.L.E., $\boldsymbol{W}/n$, and also dominates $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}$ with respect to Stein's loss. See [4,22] for more details. Order preservation among $\psi_i(\boldsymbol{l}),\ i = 1, \ldots, p$, is discussed in [19].

The orthogonally equivariant estimator, $\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$, defined by

$$\psi_i(\boldsymbol{l}) = l_i \delta_i^{\mathrm{OS}}, \quad 1 \leqslant i \leqslant p,$$

with δ_i^{OS} in (17) has been considered to be minimax from an analogy between $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ (see [9]). For the case $p = 2$, this conjecture was proved by Sheena [18].

Let $r = r(\boldsymbol{l})$ be defined in (3). For the rest of this section, *we fix an arbitrary* $\boldsymbol{\Gamma} \in \mathcal{O}(p)$ *and consider the limit* $\rho = \rho(\boldsymbol{\Lambda}) \to 0$ *in* $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$. This is the same setup as in Lemma 2. Correspondingly, in view of Corollary 1 and Theorem 1, we consider behavior of the estimators when $r(\boldsymbol{l})$ is small and $\boldsymbol{G}$ is close to $\boldsymbol{\Gamma}$. The reason for this setup is that as $\rho \to 0$, $\boldsymbol{G}$ and r converge respectively to $\boldsymbol{\Gamma}$ and 0 in probability and hence the risk function of an estimator should depend only on $\boldsymbol{G}$ close to $\boldsymbol{\Gamma}$ and small r.

Now we introduce some regularity conditions on $\widehat{\boldsymbol{\Sigma}}$, which exclude pathological cases.

Assumption 1. 1. *There exist* $0 < \bar{c}_i(\boldsymbol{\Gamma}) < \infty$, $1 \leqslant i \leqslant p$, *and* $\bar{\boldsymbol{H}}(\boldsymbol{\Gamma}) \in \mathcal{O}(p)$ *such that as* $\boldsymbol{G} \to \boldsymbol{\Gamma}$ *and* $r \to 0$,

$$c_i(\boldsymbol{G}, \boldsymbol{l}) \to \bar{c}_i(\boldsymbol{\Gamma}), \quad 1 \leqslant i \leqslant p, \quad \boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}) \to \bar{\boldsymbol{H}}(\boldsymbol{\Gamma}). \tag{20}$$

2. *There exists* $M = M(\boldsymbol{\Gamma}) < \infty$ *such that*

$$E[\operatorname{tr} \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{\Sigma}^{-1}] < M \tag{21}$$

for all $\boldsymbol{\Lambda}$ *in* $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$.

The first part of the assumption just requires that the components of an estimator converge somewhere. As for the second part, note that $\operatorname{tr} \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{\Sigma}^{-1}$ is a component of both Stein's loss function and the quadratic loss function. Therefore, estimators with bounded risks under either loss function are supposed to satisfy (21). Note that a minimax estimator has a bounded risk by definition.

Under these assumptions we have the following results.

Lemma 3. *Let* $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) = \boldsymbol{H}(\boldsymbol{W})\boldsymbol{D}(\boldsymbol{W})\boldsymbol{H}'(\boldsymbol{W})$ *be an estimator of* $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$ *satisfying Assumption* 1. *Then as* $\rho = \rho(\boldsymbol{\Lambda}) \to 0$,

$$\boldsymbol{H}(\boldsymbol{W}) \xrightarrow{p} \boldsymbol{\Gamma}.$$

Theorem 3. *Let* $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})$ *be an estimator satisfying Assumption* 1. *Then* $\bar{\boldsymbol{H}}(\boldsymbol{\Gamma}) = \boldsymbol{\Gamma}$.

Proof. Consider the triangular inequality

$$d(\boldsymbol{\Gamma}, \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})) \leqslant d(\boldsymbol{\Gamma}, \boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l})) + d(\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}), \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})). \tag{22}$$

By Lemma 3, $d(\boldsymbol{\Gamma}, \boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l})) \xrightarrow{p} 0$ as $\rho \to 0$. By Corollary 1 and Theorem 1, $r \xrightarrow{p} 0$ and $\boldsymbol{G} \xrightarrow{p} \boldsymbol{\Gamma}$ as $\rho \to 0$. Therefore by Assumption 1, $d(\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}), \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})) \xrightarrow{p} 0$. It follows that the right-hand side of (22) converges to 0 in probability as $\rho \to 0$. However, the left-hand side is a constant and therefore $d(\boldsymbol{\Gamma}, \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})) = 0$. $\square$

Let $\widetilde{\boldsymbol{H}}(\boldsymbol{G}, \boldsymbol{l}) = (\widetilde{h}_{ij}(\boldsymbol{G}, \boldsymbol{l})) = \boldsymbol{\Gamma}'\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l})$. Theorem 3 says that for a reasonable estimator, $\widetilde{\boldsymbol{H}}(\boldsymbol{G}, \boldsymbol{l}) \to \boldsymbol{I_p}$ as $\boldsymbol{G} \to \boldsymbol{\Gamma}$ and $r \to 0$; i.e., $\widetilde{h}_{ij}(\boldsymbol{G}, \boldsymbol{l}) \to 0,\ 1 \leqslant \forall j < \forall i \leqslant p$. Actually from the proof of Lemma 3 we see that $\tilde{h}_{ij}$ is of the same order as $\tilde{g}_{ij}$, i.e., $\tilde{h}_{ij} = O_p(\lambda_i^{\frac{1}{2}}/\lambda_j^{\frac{1}{2}})$. Let

$$\zeta_{ij}(\boldsymbol{G}, \boldsymbol{l}) = \frac{\widetilde{h}_{ij}(\boldsymbol{G}, \boldsymbol{l})}{\widetilde{g}_{ij}(\boldsymbol{G})}.$$

In order to evaluate the asymptotic risk of an estimator, we need to know the limit of ζ_{ij} as $\boldsymbol{G} \to \boldsymbol{\Gamma}$ and $r \to 0$ for $1 \leqslant j < i \leqslant p$. Let $\zeta_{ij,\boldsymbol{\Gamma}}$, $\widetilde{h}_{ij,\boldsymbol{\Gamma}}$ and $\widetilde{g}_{ij,\boldsymbol{\Gamma}}$ denote the local expressions around $\boldsymbol{\Gamma}$ of ζ_{ij}, $\widetilde{h}_{ij}$ and $\widetilde{g}_{ij}$, respectively. Then

$$\zeta_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = \frac{\widetilde{h}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})}{\widetilde{g}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})} = \frac{\widetilde{h}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})}{u_{ij}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})} = \frac{\widetilde{h}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})}{q_{ij} f_j^{-\frac{1}{2}} \lambda_j^{-\frac{1}{2}} \lambda_i^{\frac{1}{2}}}.$$

We now assume the existence of a limit of $\zeta_{ij,\boldsymbol{\Gamma}}$.

Assumption 2. *There exist* $\bar{\zeta}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}),\ 1 \leqslant j < i \leqslant p$, *such that*

$$\zeta_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) \to \bar{\zeta}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}) \quad a.e.\ in\ (\boldsymbol{f}, \boldsymbol{q})\ as\ \rho(\boldsymbol{\Lambda}) \to 0.$$

Note that this assumption implies $\widetilde{h}_{ij,\boldsymbol{\Gamma}} \to 0$ as $\rho \to 0$ for $1 \leqslant \forall j < \forall i \leqslant p$. The following theorem gives the asymptotic lower bounds of the risks, $R_i(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = E[L_i(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$, $i = 1, 2$.

Theorem 4. *Under Assumptions 1 and 2*

$$\varliminf_{\rho \to 0} R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Gamma\Lambda\Gamma}') \geqslant \sum_{i=1}^{p}\sum_{j=1}^{p} \bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ij}^2] - \sum_{i=1}^{p} \log \bar{c}_i(\boldsymbol{\Gamma}) - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}] - p, \tag{23}$$

$$\varliminf_{\rho \to 0} R_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Gamma\Lambda\Gamma}') \geqslant \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}] - 2\sum_{i=1}^{p}\sum_{j=1}^{p} \bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ij}^2] + p, \tag{24}$$

where

$$\xi_{ij}(\boldsymbol{f}, \boldsymbol{q}) = \begin{cases} \bar{\zeta}_{ij,\boldsymbol{\Gamma}}\, q_{ij} & \text{if } i > j, \\ f_i^{\frac{1}{2}} & \text{if } i = j, \\ 0 & \text{if } i < j \end{cases}$$

and the expectation on the right-hand side is taken with respect to the asymptotic distribution of $(\boldsymbol{f}, \boldsymbol{q})$ *given in Theorem* 2.

If in addition

$$\exists\, a < \tfrac{1}{8},\ \exists b > 0,\ 1 \leqslant \forall i, \forall j \leqslant p, \quad |\zeta_{ij}(\boldsymbol{\Gamma G}, \boldsymbol{l}) \leqslant b\ \exp\{\mathrm{tr}(a\boldsymbol{GLG}'\boldsymbol{\Lambda}^{-1})\}| \quad a.e.\ in\ (\boldsymbol{G}, \boldsymbol{l}) \tag{25}$$

and

$$0 < \exists c_l < \exists c_u < \infty,\ 1 \leqslant \forall i \leqslant p, \quad c_l < c_i(\boldsymbol{W}) < c_u, \tag{26}$$

then $\lim_{\rho\to 0} R_i(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Gamma\Lambda\Gamma}')$ $(i = 1, 2)$ *exist and equal the right-hand side of* (23) $(i = 1)$ *or* (24) $(i = 2)$.

Proof. Here for notational simplicity we prove the theorem for the case of $\boldsymbol{\Gamma} = \boldsymbol{I_p}$. This can be done without loss of generality, because given an estimator $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})$ and $\boldsymbol{\Gamma}$, we can consider an estimator $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}'\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Gamma W \Gamma}')\boldsymbol{\Gamma}$ instead of $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})$. Therefore, in this proof we simply write $\boldsymbol{G}$ instead of $\widetilde{\boldsymbol{G}}$. Similarly we simply denote the local expression of a function $x(\boldsymbol{G}, \boldsymbol{\lambda})$ around $\boldsymbol{I_p}$ by $x(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})$.

Using the notation $\alpha_{ij} = \alpha_{ij}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) = h_{ij}(\boldsymbol{G}, \boldsymbol{l})c_j^{\frac{1}{2}} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}},\ 1 \leqslant i, j \leqslant p$, we can write the loss functions as

$$\begin{aligned}
L_1(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda}) &= \mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{HDH}') - \sum_{i=1}^{p} \log\, c_i - \log\, |\boldsymbol{\Lambda}^{-1}\boldsymbol{W}| - p \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{ij}^2 - \sum_{i=1}^{p} \log\, c_i - \log\, |\boldsymbol{\Lambda}^{-1}\boldsymbol{W}| - p. \\
L_2(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda}) &= \mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{HDH}' - \boldsymbol{I_p})^2 \\
&= \mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{HDH}')^2 - 2\,\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{HDH}') + p \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \alpha_{ki}\alpha_{kj}\alpha_{li}\alpha_{lj} - 2\sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{ij}^2 + p.
\end{aligned}$$

The proof proceeds similarly as the one in Lemma 2. For $i = 1, 2$, we have

$$\begin{aligned}
R_i(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda}) &= E[I((\boldsymbol{l}, \boldsymbol{G}) \in \mathcal{M}(\boldsymbol{I_p}))L_i(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda})] + E[I((\boldsymbol{l}, \boldsymbol{G}) \in \mathcal{M}(\boldsymbol{I_p})^C)L_i(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda})] \\
&\geqslant E[I((\boldsymbol{l}, \boldsymbol{G}) \in \mathcal{M}(\boldsymbol{I_p}))L_i(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda})] \\
&= \int_{R_+^p}\int_{R^{\frac{p(p-1)}{2}}} L_i(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\, d\boldsymbol{q}\, d\boldsymbol{f}.
\end{aligned}$$

By Fatou's lemma, we have

$$\begin{aligned}
\underline{\lim}_{\rho\to 0}\, R_i(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda}) &\geqslant \int_{R_+^p}\int_{R^{\frac{p(p-1)}{2}}} \underline{\lim}_{\rho\to 0}\, L_i(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\, d\boldsymbol{q}\, d\boldsymbol{f} \\
&= \int_{R_+^p}\int_{R^{\frac{p(p-1)}{2}}} \lim_{\rho\to 0}\, L_i(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\, d\boldsymbol{q}\, d\boldsymbol{f}.
\end{aligned}$$

Using the fact $E[\log|\boldsymbol{\Lambda}^{-1}\boldsymbol{W}|] = \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}]$, we have

$$L_1(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{ij}^2 - \sum_{j=1}^{p} \log c_j - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}] - p$$

around $\boldsymbol{I_p}$. Note that $\alpha_{ij}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = h_{ij}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) c_j^{\frac{1}{2}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) f_j^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}$ around $\boldsymbol{I_p}$. By Assumption 1 we have

$$\lim_{\rho\to 0} c_j(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = \lim_{\rho\to 0} c_j(\boldsymbol{G}(\boldsymbol{u}), \boldsymbol{l}(\boldsymbol{f}, \boldsymbol{\lambda})) = \lim_{G\to \boldsymbol{I_p}, r\to 0} c_j(\boldsymbol{G}, \boldsymbol{l}) = \bar{c}_j(\boldsymbol{I_p}).$$

Using this convergence, Assumption 2 and the fact $h_{ii} \to 1,\ 1 \leqslant \forall i \leqslant p,\ h_{ij} \to 0,\ 1 \leqslant \forall i < \forall j \leqslant p$, we have

$$\lim_{\rho\to 0} \alpha_{ij} = \begin{cases} \lim_{\rho\to 0} \zeta_{ij} c_j^{\frac{1}{2}} q_{ij} = \bar{\zeta}_{ij} \bar{c}_j(\boldsymbol{I_p})^{\frac{1}{2}} q_{ij} & \text{if } i > j, \\ \lim_{\rho\to 0} h_{ii} c_i^{\frac{1}{2}} f_i^{\frac{1}{2}} = \bar{c}_i(\boldsymbol{I_p})^{\frac{1}{2}} f_i^{\frac{1}{2}} & \text{if } i = j, \\ 0 & \text{if } i < j. \end{cases}$$

This can be written uniformly as $\lim_{\rho\to 0} \alpha_{ij} = \bar{c}_j(\boldsymbol{I_p})^{\frac{1}{2}} \xi_{ij}$. Therefore,

$$\lim_{\rho\to 0} \log c_j = \log \bar{c}_j(\boldsymbol{I_p}),$$

$$\lim_{\rho\to 0} \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{ij}^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} \bar{c}_j(\boldsymbol{I_p}) \xi_{ij}^2,$$

$$\lim_{\rho\to 0} \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \alpha_{ki}\alpha_{kj}\alpha_{li}\alpha_{lj} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \bar{c}_i(\boldsymbol{I_p})\bar{c}_j(\boldsymbol{I_p}) \xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}.$$

Since $h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})$ converges to the density of the asymptotic distributions given in Theorem 2, the inequalities in the lemma are proved.

Now we assume (25) and (26). Since $|\alpha_{ij}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})| \leqslant c_u^{\frac{1}{2}} |\zeta_{ij}(\boldsymbol{G}, \boldsymbol{l})||g_{ij}| l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}$,

$$\begin{aligned} &\exists\, a' < 1/8 - a,\ \exists b' > 0,\ 1 \leqslant \forall i, \forall j \leqslant p, \\ &\quad |\alpha_{ij}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})| \leqslant |\zeta_{ij}(\boldsymbol{G}, \boldsymbol{l})|\, b' \mathrm{etr}(a' \boldsymbol{GLG}' \boldsymbol{\Lambda}^{-1}). \end{aligned}$$

Therefore for $1 \leqslant \forall i, \forall j, \forall k, \forall l \leqslant p$,

$$\alpha_{ij}^2(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) \leqslant (bb')^2 \mathrm{etr}\{2(a + a')\boldsymbol{GLG}'\boldsymbol{\Lambda}^{-1}\} \quad \text{a.e.},$$

$$\begin{aligned} &|\alpha_{ki}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})\alpha_{kj}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})\alpha_{li}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})\alpha_{lj}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})| \\ &\quad \leqslant (bb')^4 \mathrm{etr}\{4(a + a')\boldsymbol{GLG}'\boldsymbol{\Lambda}^{-1}\} \quad \text{a.e.}, \end{aligned}$$

$$|\log c_i(\boldsymbol{G}, \boldsymbol{l})| \leqslant \max(|\log c_l|, |\log c_u|).$$

By Lemma 2, for $i = 1, 2$,

$$\lim_{\rho\to 0} R(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda}) = E\left[\lim_{\rho\to 0} L_i(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\right]. \qquad \square$$

Note that if we have the following inequality in the sense of non-negative definiteness

$$c_l \boldsymbol{W} \leqslant \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) \leqslant c_u \boldsymbol{W}, \tag{27}$$

then $c_l < c_i(\boldsymbol{W}) < c_u$, $i = 1 \leqslant \forall i \leqslant p$, by the minimax characterization of the eigenvalues (see for example [11, Chapter 20 A.1.b]). In particular (27) holds for triangularly equivariant estimators $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{T}\,\mathrm{diag}(\delta_1, \ldots, \delta_p)\boldsymbol{T}'$ with $c_l = \min(\delta_1, \ldots, \delta_p)$ and $c_u = \max(\delta_1, \ldots, \delta_p)$.

Theorem 4 shows that the values of $\bar{\zeta}_{ij,\boldsymbol{\Gamma}}$, $1 \leqslant j < i \leqslant p$, determine the risk or its lower bound.

Now we state a corollary on necessary conditions for $\widehat{\boldsymbol{\Sigma}}$ to be tail minimax.

Corollary 2. *Suppose that $\widehat{\boldsymbol{\Sigma}}$ satisfies Assumptions 1 and 2 with*

$$\bar{\zeta}_{ij,\boldsymbol{\Gamma}} = 1, \quad 1 \leqslant \forall j < \forall i \leqslant p. \tag{28}$$

If it is tail minimax with respect to Stein's loss, then

$$\bar{c}_i(\boldsymbol{\Gamma}) = \delta_i^{\mathrm{JS}}, \quad i = 1, \ldots, p,$$

where δ_i^{JS} is given in (13). If it is tail minimax with respect to the quadratic loss, then

$$\bar{c}_i(\boldsymbol{\Gamma}) = \delta_i^{\mathrm{OS}}, \quad i = 1, \ldots, p,$$

where δ_i^{OS} is given in (17).

Proof. We first consider Stein's loss. Substitute

$$\xi_{ij}(\boldsymbol{f}, \boldsymbol{q}) = \begin{cases} q_{ij} & \text{if } i > j, \\ f_i^{\frac{1}{2}} & \text{if } i = j, \\ 0 & \text{if } i < j \end{cases}$$

into (23). Then we have

$$\begin{aligned} &\varliminf_{\rho\to 0} R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) \\ &\quad \geqslant \sum_{i>j} \bar{c}_j(\boldsymbol{\Gamma}) E[\chi_1^2] + \sum_{i=1}^{p} \bar{c}_i(\boldsymbol{\Gamma}) E[\chi^2_{n-i+1}] - \sum_{i=1}^{p} \log\, \bar{c}_i(\boldsymbol{\Gamma}) \\ &\qquad - \sum_{i=1}^{p} E[\log\, \chi^2_{n-i+1}] - p \end{aligned}$$

$$= \sum_{i>j} \bar{c}_j(\boldsymbol{\Gamma}) + \sum_{i=1}^{p} \bar{c}_i(\boldsymbol{\Gamma})(n-i+1) - \sum_{i=1}^{p} \log \bar{c}_i(\boldsymbol{\Gamma}) - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}] - p$$
$$= \sum_{i=1}^{p} \{\bar{c}_j(\boldsymbol{\Gamma})/\delta_i^{\mathrm{JS}} - \log \bar{c}_j(\boldsymbol{\Gamma})\} - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}] - p$$

with δ_i^{JS} given in (13). The right side is uniquely minimized when $\bar{c}_i(\boldsymbol{\Gamma}) = \delta_i^{\mathrm{JS}}$. Therefore,

$$\underline{\lim}_{\rho\to 0}\, R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) \geqslant \sum_{i=1}^{p} \log(n+p+1-2i) - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}],$$

where the right side is equal to the minimax risk $\bar{R}_1$ in (14). If $\bar{c}_i(\boldsymbol{\Gamma}) \neq \delta_i^{\mathrm{JS}},\ 1 \leqslant \exists i \leqslant p$, then for any small δ there exists some $\boldsymbol{\Sigma} = \boldsymbol{\Gamma\Lambda\Gamma}'$ such that

$$\rho(\boldsymbol{\Lambda}) < \delta, \quad R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) > \bar{R}_1.$$

But this contradicts the fact that $\widehat{\boldsymbol{\Sigma}}$ is a tail minimax estimator. Consequently, it is necessary that

$$\bar{c}_i(\boldsymbol{\Gamma}) = \delta_i^{\mathrm{JS}}, \quad 1 \leqslant \forall i \leqslant p.$$

This completes the proof for Stein's loss.

Next we give a proof for the quadratic loss. The proof is straightforward but long. We briefly sketch the outline. We decompose the four-folded summation in (23) as

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}]$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ki}^2\xi_{kj}^2] + 2\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k<l} \bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}].$$

Furthermore,

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ki}^2\xi_{kj}^2]$$
$$= \sum_{k<i \text{ or } k<j} + \sum_{k=i=j} + \sum_{k=i,k>j} + \sum_{k=j,k>i} + \sum_{k>i,k>j,i=j} + \sum_{k>i,k>j,i\neq j}$$
$$= \sum_{i=1}^{p} \{(n-i+1)(n-i+3) + 3(p-i)\}\bar{c}_i(\boldsymbol{\Gamma})^2$$
$$+ \sum_{i<j} (n+p-2j+1)\bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma})$$
$$+ \sum_{j<i} (n+p-2i+1)\bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}).$$

We also have

$$
\begin{aligned}
&2\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k<l}\bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma})E[\xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}]\\
&\quad = 2\sum_{i=1}^{p}\sum_{k<l}\bar{c}_i^2(\boldsymbol{\Gamma})E[\xi_{ki}^2\xi_{li}^2] + 4\sum_{i<j}\sum_{k<l}\bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma})E[\xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}]\\
&\quad = \sum_{i=1}^{p}\bar{c}_i^2(\Gamma)\{2(p-i)(n-i+1)+(p-i)(p-i-1)\}.
\end{aligned}
$$

Combining these results and the result on $R_1(\widehat{\Sigma}, \Sigma)$,

$$
2\sum_{i=1}^{p}\sum_{j=1}^{p}\bar{c}_j(\boldsymbol{\Gamma})E[\xi_{ij}^2] = 2\sum_{i=1}^{p}(n+p+1-2i)\bar{c}_i(\boldsymbol{\Gamma}),
$$

we obtain

$$
\varliminf_{\rho\to 0} R_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Gamma\Lambda\Gamma}') \geq \boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} - 2\boldsymbol{b}'\boldsymbol{c} + p, \tag{29}
$$

where $\boldsymbol{c} = (\bar{c}_1(\boldsymbol{\Gamma}), \ldots, \bar{c}_p(\boldsymbol{\Gamma}))'$ and the elements $\boldsymbol{A}$ and $\boldsymbol{b}$ are given in (15) and (16). The minimum of the quadratic function (29) is uniquely attained when $\boldsymbol{c}$ satisfies the linear equation $\boldsymbol{Ac} = \boldsymbol{b}$, namely

$$
\bar{c}_i(\boldsymbol{\Gamma}) = \delta_i^{\mathrm{OS}}, \quad 1 \leqslant \forall i \leqslant p,
$$

where δ_i^{OS} is defined in (17) and the minimum value of the quadratic function is the minimax risk $\bar{R}_2$ in (18). This proves the theorem for the case of the quadratic loss. □

Note that Corollary 2 holds for any orthogonally equivariant estimator in (19), since $\widetilde{h}_{ij} = \widetilde{g}_{ij}$ and

$$
\zeta_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) \equiv 1, \quad 1 \leqslant \forall j < \forall i \leqslant p,
$$

for orthogonally equivariant estimators. In view of this, (28) seems to imply that we are restricted to estimators which are nearly orthogonally equivariant. However, we conjecture (28) should hold a.e. in $\boldsymbol{\Gamma}$ for any tail minimax estimator. An argument on this point is given in the preprint version [23] of this paper, available from the authors.

Roughly speaking, Theorem 3 and Corollary 2 indicate that when r is very small, any tail minimax estimator satisfying Assumptions 1, 2 with (28) must be approximately the same as $\widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS}}$ in the case of Stein's loss and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{KG}}$ in the case of the quadratic loss, respectively; i.e.,

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}(\boldsymbol{G}, \boldsymbol{l}) &= \boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l})\,\mathrm{diag}(l_1 c_1(\boldsymbol{G}, \boldsymbol{l}), \ldots, l_p c_p(\boldsymbol{G}, \boldsymbol{l}))\boldsymbol{H}'(\boldsymbol{G}, \boldsymbol{l})\\
&\approx \boldsymbol{G}\,\mathrm{diag}(l_1\delta_1^{\mathrm{JS(OS)}}, \ldots, l_p\delta_p^{\mathrm{JS(OS)}})\boldsymbol{G}'\\
&= \widehat{\boldsymbol{\Sigma}}^{\mathrm{SDS(KG)}}.
\end{aligned}
$$

We state one simple application of Corollary 2. The orthogonally equivariant estimators contain the subclass (say C^o) of estimators which is defined by

$$c_i(\boldsymbol{l}) = c_i \text{ (constant)}, \quad 1 \leqslant i \leqslant p.$$

This class contains the M.L.E., $\boldsymbol{W}/n$ and $\widehat{\boldsymbol{\Sigma}}^{\text{SDS}}, \widehat{\boldsymbol{\Sigma}}^{\text{OS}}$. We have conjectured that $\widehat{\boldsymbol{\Sigma}}^{\text{SDS}}$ and $\widehat{\boldsymbol{\Sigma}}^{\text{OS}}$ are the only minimax estimators in this class with respect to Stein's loss and the quadratic loss. This conjecture was proved in Sheena [18] for the case $p = 2$. It is obvious that every estimator in C^o satisfies Assumptions 1 and 2. Then from Corollary 2, it is necessary that

$$c_i = \delta_i^{\text{JS(OS)}}, \quad 1 \leqslant i \leqslant p,$$

which shows that the above conjecture holds true for general dimension.

Acknowledgements

We thank the referees for their helpful suggestions which greatly improved the presentation of this paper. The second author is deeply grateful to the Department of Mathematics and Statistics, Bowling Green State University for providing comfortable research facilities.

AppendixA. Proof of Lemmas

Proof of Lemma 1. Since $\boldsymbol{L}(\boldsymbol{\Gamma}'\boldsymbol{W}\boldsymbol{\Gamma}) = \boldsymbol{L}(\boldsymbol{W})$ and $\boldsymbol{\Gamma}'\boldsymbol{W}\boldsymbol{\Gamma} \sim \boldsymbol{W}_p(n, \boldsymbol{\Lambda})$, it suffices to show

$$P\left(C_1 < \frac{l_i}{\lambda_i} < C_2,\ 1 \leqslant \forall i \leqslant p\right) > 1 - \varepsilon \quad \forall \boldsymbol{\Lambda}.$$

Let

$$\widetilde{\boldsymbol{W}} = (\widetilde{w}_{ij}) = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Lambda}^{-\frac{1}{2}} \sim \boldsymbol{W}_p(n, \boldsymbol{I_p}).$$

First we consider l_1. Since

$$l_1 \leqslant \operatorname{tr} \boldsymbol{W} = w_{11} + \cdots + w_{pp} = \lambda_1 \widetilde{w}_{11} + \cdots + \lambda_p \widetilde{w}_{pp},$$

we have

$$\frac{l_1}{\lambda_1} \leqslant \widetilde{w}_{11} + \frac{\lambda_2}{\lambda_1}\widetilde{w}_{22} + \cdots + \frac{\lambda_p}{\lambda_1}\widetilde{w}_{pp} \leqslant \widetilde{w}_{11} + \widetilde{w}_{22} + \cdots + \widetilde{w}_{pp} \overset{d}{=} \chi^2_{np}.$$

Hence

$$P\left(\frac{l_1}{\lambda_1} \geqslant C_2\right) \leqslant P\left(\chi^2_{np} \geqslant C_2\right) \quad \forall C_2 > 0. \tag{A.1}$$

This means

$$P\left(\frac{l_1}{\lambda_1} \geqslant C_2\right) \to 0 \quad \text{as } C_2 \to \infty$$

uniformly in $\boldsymbol{\Lambda}$.

Next we consider l_2. Let $\boldsymbol{W}_{22}$ be the $(p-1)\times(p-1)$ matrix that is made by deleting the first column and the first row of $\boldsymbol{W}$. Then $\boldsymbol{W}_{22} \sim \boldsymbol{W}_{p-1}(n, \boldsymbol{\Lambda}_{22})$ with $\boldsymbol{\Lambda}_{22} = \mathrm{diag}(\lambda_2, \ldots, \lambda_p)$. Let $\tilde{l}_2$ be the largest eigenvalue of $\boldsymbol{W}_{22}$. Then by the minimax characterization of eigenvalues (see for example [11, Chapter 20 A.1.c])

$$l_2 \leqslant \tilde{l}_2.$$

If we use the result (A.1) for $\tilde{l}_2/\lambda_2$, we have

$$P\left(\frac{l_2}{\lambda_2} \geqslant C_2\right) \leqslant P\left(\frac{\tilde{l}_2}{\lambda_2} \geqslant C_2\right) \leqslant P(\chi^2_{n(p-1)} \geqslant C_2) \quad \forall C_2 > 0.$$

Therefore

$$P\left(\frac{l_2}{\lambda_2} \geqslant C_2\right) \to 0 \quad \text{as } C_2 \to \infty,$$

uniformly in $\boldsymbol{\Lambda}$. Completely similarly we can prove

$$P\left(\frac{l_i}{\lambda_i} \geqslant C_2\right) \to 0 \quad \text{as } C_2 \to \infty, \quad 1 \leqslant \forall i \leqslant p, \tag{A.2}$$

uniformly in $\boldsymbol{\Lambda}$.

Now we consider the reverse inequality. We use the inverse Wishart distribution. Let

$$\widetilde{\boldsymbol{W}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Lambda}^{-\frac{1}{2}}, \quad \widetilde{\boldsymbol{W}}^{-1} = (\widetilde{w}^{ij}) = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{W}^{-1}\boldsymbol{\Lambda}^{\frac{1}{2}}.$$

Then $\widetilde{\boldsymbol{W}}^{-1} \sim \boldsymbol{W}_p^{-1}(n, \boldsymbol{I_p})$ and its distribution is independent of $\boldsymbol{\Lambda}$. First we consider l_p^{-1}, the largest eigenvalue of $\boldsymbol{W}^{-1}$. Since

$$\frac{1}{l_p} \leqslant \mathrm{tr}\, \boldsymbol{W}^{-1} = \frac{\widetilde{w}^{11}}{\lambda_1} + \cdots + \frac{\widetilde{w}^{pp}}{\lambda_p} \leqslant \frac{1}{\lambda_p}(\widetilde{w}^{11} + \cdots + \widetilde{w}^{pp}),$$

we have

$$P\left(\frac{\lambda_p}{l_p} \geqslant \frac{1}{C_1}\right) \leqslant P\left(\widetilde{w}^{11} + \cdots + \widetilde{w}^{pp} \geqslant \frac{1}{C_1}\right) \quad \forall C_1 > 0.$$

The right side is independent of $\boldsymbol{\Lambda}$. Hence

$$P\left(\frac{l_p}{\lambda_p} \leqslant C_1\right) \to 0 \quad \text{as } C_1 \to 0,$$

uniformly in $\boldsymbol{\Lambda}$.

Next we consider l_{p-1}^{-1}. Let $\boldsymbol{W}_{(2)}^{-1}$ be the $(p-1)\times(p-1)$ matrix made by deleting the last column and the last row of $\boldsymbol{W}^{-1}$. Since $\boldsymbol{W}^{-1} \sim \boldsymbol{W}_p^{-1}(n, \boldsymbol{\Lambda}^{-1})$, we have

$$\boldsymbol{W}_{(2)}^{-1} \sim \boldsymbol{W}_{p-1}^{-1}(n-1, \boldsymbol{\Lambda}^{11}), \quad \boldsymbol{\Lambda}^{11} = \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_{p-1}^{-1}).$$

Let l_2^* denote the largest eigenvalue of $\boldsymbol{W}_{(2)}^{-1}$. Then

$$\frac{1}{l_{p-1}} \leqslant l_2^*.$$

We also have

$$l_2^* \leqslant \operatorname{tr} \boldsymbol{W}_{(2)}^{-1} = \frac{\widetilde{w}^{11}}{\lambda_1} + \cdots + \frac{\widetilde{w}^{p-1,p-1}}{\lambda_{p-1}} \leqslant \frac{1}{\lambda_{p-1}}(\widetilde{w}^{11} + \cdots + \widetilde{w}^{p-1,p-1}).$$

Therefore,

$$P\left(\frac{\lambda_{p-1}}{l_{p-1}} \geqslant \frac{1}{C_1}\right) \leqslant P\left(\lambda_{p-1} l_2^* \geqslant \frac{1}{C_1}\right) \leqslant P\left(\widetilde{w}^{11} + \cdots + \widetilde{w}^{p-1,p-1} \geqslant \frac{1}{C_1}\right) \quad \forall C_1 > 0$$

and

$$P\left(\frac{l_{p-1}}{\lambda_{p-1}} \leqslant C_1\right) \to 0 \quad \text{as } C_1 \to 0,$$

uniformly in $\boldsymbol{\Lambda}$. Completely similarly we can prove

$$P\left(\frac{l_i}{\lambda_i} \leqslant C_1\right) \to 0 \quad \text{as } C_1 \to 0, \quad 1 \leqslant \forall i \leqslant p, \tag{A.3}$$

uniformly in $\boldsymbol{\Lambda}$.

From (A.2), (A.3) and the Bonferroni inequality, we can choose C_1 and C_2 for any given $\varepsilon > 0$ so that

$$P\left(C_1 < \frac{l_i}{\lambda_i} < C_2,\ 1 \leqslant \forall i \leqslant p\right) > 1 - \varepsilon \quad \forall \boldsymbol{\Lambda}. \qquad \square$$

Proof of Lemma 2. The random variables $\boldsymbol{l} = (l_1, \ldots, l_p)$ and $\widetilde{\boldsymbol{G}} = \boldsymbol{\Gamma}'\boldsymbol{G}$ have the following joint density function with respect to the Lebesgue measure on R^p and the invariant probability μ on $\mathcal{O}(p)^+ = \{\widetilde{\boldsymbol{G}} \in \mathcal{O}(p) \mid \tilde{g}_{ii} \geqslant 0,\ 1 \leqslant \forall i \leqslant p\}$.

$$c_1 \prod_{i=1}^{p} \lambda_i^{-\frac{n}{2}} \prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} \prod_{j<i} (l_j - l_i) \operatorname{etr}\left(-\frac{1}{2}\widetilde{\boldsymbol{G}}\boldsymbol{L}\widetilde{\boldsymbol{G}}'\boldsymbol{\Lambda}^{-1}\right),$$

where $\operatorname{etr} X = \exp(\operatorname{tr} X)$. For the present proof we do not need an explicit form of the normalizing constant c_1. Therefore

$$\begin{aligned} E[x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})] &= E[x(\boldsymbol{\Gamma}\widetilde{\boldsymbol{G}}, \boldsymbol{l}, \boldsymbol{\lambda})] \\ &= c_1 \prod_{i=1}^{p} \lambda_i^{-\frac{n}{2}} \int_{\mathcal{L}} \int_{\mathcal{O}(p)^+} x(\boldsymbol{\Gamma}\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) \prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} \\ &\quad \times \prod_{j<i} (l_j - l_i) \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{G}\boldsymbol{L}\boldsymbol{G}'\boldsymbol{\Lambda}^{-1}\right) d\mu(\boldsymbol{G})\, d\boldsymbol{l}, \end{aligned} \tag{A.4}$$

where

$$\mathcal{L} = \{\boldsymbol{l} \mid l_1 > l_2 > \cdots > l_p > 0\}.$$

For the integration with respect to $\boldsymbol{G}$ on $\mathcal{O}(p)^+$, we consider the integration in a neighborhood of $\boldsymbol{I_p}$ and outside the neighborhood separately. We define a neighborhood of $\boldsymbol{I_p}$ using the expression $\boldsymbol{G}(\boldsymbol{u})$. Since $g_{ii}(\boldsymbol{0}) = 1,\ 1 \leqslant \forall i \leqslant p$, there exists an open set U^* that satisfies

$$\boldsymbol{0} \subset U^* \subset \bar{U}^* \subset U \quad (\bar{U}^* \text{ is the closure of } U^*)$$

and

$$g_{ii}(\boldsymbol{u}) \geqslant \sqrt{\frac{1}{2}}, \quad 1 \leqslant \forall i \leqslant p \ \forall \boldsymbol{u} \in \bar{U}^*. \tag{A.5}$$

Let

$$\mathcal{M}(\boldsymbol{I_p}) = \mathcal{L} \times \boldsymbol{G}(\bar{U}^*) = \mathcal{L} \times \{\boldsymbol{G}(\boldsymbol{u}) \mid \boldsymbol{u} \in \bar{U}^*\}.$$

Integral (A.4) is divided into two parts, say I_1 over $\mathcal{M}(\boldsymbol{I_p})$ and I_2 over $\mathcal{M}(\boldsymbol{I_p})^C = \mathcal{L} \times \boldsymbol{G}(\bar{U}^*)^C$. Then from (11)

$$\begin{aligned} I_2 &\leqslant c_1 \prod_{i=1}^p \lambda_i^{-\frac{n}{2}} \int_{\mathcal{M}(I_p)^C} |x(\boldsymbol{\Gamma G}, \boldsymbol{l}, \boldsymbol{\lambda})| \prod_{i=1}^p l_i^{\frac{n-p-1}{2}} \\ &\quad \times \prod_{j<i} (l_j - l_i) \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{GLG}'\boldsymbol{\Lambda}^{-1}\right) d\mu(\boldsymbol{G})\, d\boldsymbol{l} \\ &\leqslant c_1 b \prod_{i=1}^p \lambda_i^{-\frac{n}{2}} \int_{\mathcal{M}(I_p)^C} \prod_{i=1}^p l_i^{\frac{n-p-1}{2}} \prod_{j<i} (l_j - l_i) \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{GLG}'\widetilde{\boldsymbol{\Lambda}}^{-1}\right) d\mu(\boldsymbol{G})\, d\boldsymbol{l} \\ &= c_1' P(\boldsymbol{\Gamma}'\boldsymbol{G} \notin \boldsymbol{G}(\bar{U}^*) \mid \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma}'), \end{aligned} \tag{A.6}$$

where $\widetilde{\boldsymbol{\Lambda}} = (1-2a)^{-1}\boldsymbol{\Lambda}$. Since $P(\boldsymbol{\Gamma}'\boldsymbol{G} \notin \boldsymbol{G}(\bar{U}^*) \mid \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma}') \to 0$ as $\rho \to 0$ by Theorem 1, I_2 vanishes.

Now we focus ourselves on I_1.

$$\begin{aligned} I_1 = c_1 \prod_{i=1}^p \lambda_i^{-\frac{n}{2}} \int_{\mathcal{M}(I_p)} x(\boldsymbol{\Gamma G}, \boldsymbol{l}, \boldsymbol{\lambda}) \prod_{i=1}^p l_i^{\frac{n-p-1}{2}} \\ \times \prod_{j<i} (l_j - l_i) \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{GLG}'\boldsymbol{\Lambda}^{-1}\right) d\mu(\boldsymbol{G})\, d\boldsymbol{l}. \end{aligned}$$

We want to express the integral with respect to $d\mu(\boldsymbol{G})$ in terms of the local coordinates $\boldsymbol{u}$. It is well known that the invariant measure $d\mu(\boldsymbol{G})$ has the exterior differential form expression

$$c_2 \bigwedge_{i>j} \boldsymbol{g}_j' \, d\boldsymbol{g}_i, \tag{A.7}$$

where $\boldsymbol{g}_i$ is the ith column of $\boldsymbol{G}$. Substituting the differential

$$dg_{ij} = du_{ij}, \quad i > j,$$

$$dg_{ij} = \sum_{k>l} \frac{\partial g_{ij}}{\partial u_{kl}} du_{kl}, \quad i \leqslant j,$$

into (A.7) and taking the wedge product of the terms, we see that

$$\bigwedge_{i>j} \boldsymbol{g}_j' \, d\boldsymbol{g}_i = \pm J^*(\boldsymbol{u}) \bigwedge_{i>j} du_{ij},$$

where $J^*(\boldsymbol{u})$ is the Jacobian expressing the Radon–Nikodym derivative of the measure on $\bar{U}^*$ induced from the invariant measure on $\mathcal{O}(p)$ with respect to the Lebesgue measure on $R^{\frac{p(p-1)}{2}}$. An explicit form of $J^*(\boldsymbol{u})$ for small dimension p is discussed in Appendix B. Since $J^*(\boldsymbol{u})$ is a C^∞ function on $\bar{U}^*$, it is bounded and has a finite limit as $\boldsymbol{u} \to \boldsymbol{0}$. By the above change of variables, I_1 is written as

$$I_1 = c_3 \prod_{i=1}^{p} \lambda_i^{-\frac{n}{2}} \int_{\mathcal{L}} \int_{\bar{U}^*} x(\boldsymbol{\Gamma G}(\boldsymbol{u}), \boldsymbol{l}, \boldsymbol{\lambda})$$
$$\times \prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} \prod_{j<i} (l_j - l_i) \operatorname{etr}\left(-\frac{1}{2} \boldsymbol{G}(\boldsymbol{u}) \boldsymbol{L} \boldsymbol{G}'(\boldsymbol{u}) \boldsymbol{\Lambda}^{-1}\right) J^*(\boldsymbol{u}) \, d\boldsymbol{u} \, d\boldsymbol{l}.$$

Now we consider the further coordinate transformation $(\boldsymbol{l}, \boldsymbol{u}) \to (\boldsymbol{f}, \boldsymbol{q})$, where

$$\boldsymbol{f} = (f_i)_{1 \leqslant i \leqslant p}, \quad f_i = \frac{l_i}{\lambda_i},$$

$$\boldsymbol{q} = (q_{ij})_{1 \leqslant j < i \leqslant p}, \quad q_{ij} = u_{ij} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}.$$

The Jacobian of this transformation is

$$\left| \det \left(\frac{\partial(\boldsymbol{l},\ \boldsymbol{u})}{\partial(\boldsymbol{f},\ \boldsymbol{q})} \right) \right| = \left| \det \left(\frac{\partial \boldsymbol{l}}{\partial \boldsymbol{f}} \right) \right| \left| \det \left(\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{q}} \right) \right|$$
$$= \prod_{i=1}^{p} \lambda_i \prod_{i>j} f_j^{-\frac{1}{2}} \lambda_j^{-\frac{1}{2}} \lambda_i^{\frac{1}{2}}$$
$$= \prod_{i=1}^{p} \lambda_i^{\frac{-p+2i+1}{2}} \prod_{i=1}^{p} f_i^{-\frac{p-i}{2}}.$$

Furthermore

$$\prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} = \prod_{i=1}^{p} f_i^{\frac{n-p-1}{2}} \prod_{i=1}^{p} \lambda_i^{\frac{n-p-1}{2}},$$

$$\prod_{j<i} (l_j - l_i) = \prod_{j<i} (f_j \lambda_j - f_i \lambda_i) = \prod_{j<i} f_j \lambda_j \left(1 - \frac{f_i \lambda_i}{f_j \lambda_j}\right)$$
$$= \prod_{i=1}^{p} f_i^{p-i} \prod_{i=1}^{p} \lambda_i^{p-i} \prod_{j<i} \left(1 - \frac{f_i \lambda_i}{f_j \lambda_j}\right),$$

$$
\begin{aligned}
\operatorname{tr}\ \boldsymbol{G}(\boldsymbol{u})\boldsymbol{L}\boldsymbol{G}'(\boldsymbol{u})\boldsymbol{\Lambda}^{-1} &= \sum_{i=1}^{p}\sum_{j=1}^{p} g_{ij}^2(\boldsymbol{u}) l_j \lambda_i^{-1} \\
&= \sum_{i>j} q_{ij}^2 + \sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i + \sum_{i<j} g_{ij}^2(\boldsymbol{u}) f_j \lambda_j \lambda_i^{-1},
\end{aligned}
$$

where

$$
\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = (u_{ij})_{i>j}, \quad u_{ij} = q_{ij} f_j^{-1/2} \lambda_j^{-1/2} \lambda_i^{1/2},
$$

is now a function of $\boldsymbol{f}, \boldsymbol{q}$ as well as of $\boldsymbol{\lambda}$. We also have $\boldsymbol{l} = \boldsymbol{l}(\boldsymbol{f}, \boldsymbol{\lambda}) = (f_1\lambda_1, \ldots, f_p\lambda_p)$.

Combining the above calculations, we have

$$
I_1 = \int_{R^{p(p-1)/2}} \int_{R_+^p} x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\, d\boldsymbol{f}\, d\boldsymbol{q}, \tag{A.8}
$$

where

$$
\begin{aligned}
h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = {} & c_3 I(f_1\lambda_1 > \ldots > f_p\lambda_p) I(\boldsymbol{u} \in \bar{U}^*) \\
& \times \prod_{i=1}^{p} f_i^{\frac{n-i-1}{2}} \prod_{j<i} \left(1 - \frac{f_i\lambda_i}{f_j\lambda_j}\right) \exp\left(-\frac{1}{2}\sum_{i>j} q_{ij}^2\right) \\
& \times \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i + \sum_{i<j} g_{ij}^2(\boldsymbol{u}) f_j \lambda_j \lambda_i^{-1}\right)\right\} J^*(\boldsymbol{u}).
\end{aligned}
$$

We will show that $x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})$ is bounded in $\boldsymbol{\lambda}$. First $J^*(\boldsymbol{u})\ I(\boldsymbol{u}) \leqslant K$ for some $K > 0$, because $J^*(\boldsymbol{u})$ is bounded on a compact set $\bar{U}^*$ as remarked above. Clearly

$$
0 < \prod_{j<i} \left(1 - \frac{f_i\lambda_i}{f_j\lambda_j}\right) I(f_1\lambda_1 > \cdots > f_p\lambda_p) < 1.
$$

From (11) we have

$$
\begin{aligned}
|x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})| &= |x(\boldsymbol{\Gamma}\boldsymbol{G}(\boldsymbol{u}), \boldsymbol{l}(\boldsymbol{f}, \boldsymbol{\lambda}), \boldsymbol{\lambda})| \\
&\leqslant b\ \exp\left\{a\left(\sum_{i>j} q_{ij}^2 + \sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i + \sum_{i<j} g_{ij}^2(\boldsymbol{u}) f_j \lambda_j \lambda_i^{-1}\right)\right\}.
\end{aligned}
$$

Therefore, using (A.5), we have

$$
\begin{aligned}
& |x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})| \exp\left\{-\frac{1}{2}\left(\sum_{i>j} q_{ij}^2 + \sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i + \sum_{i<j} g_{ij}^2(\boldsymbol{u}) f_j \lambda_j \lambda_i^{-1}\right)\right\} I(\boldsymbol{u} \in \bar{U}^*) \\
& \quad \leqslant b \exp\left\{-\frac{1-2a}{2}\left(\sum_{i>j} q_{ij}^2 + \sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i + \sum_{i<j} g_{ij}^2(\boldsymbol{u}) f_j \lambda_j \lambda_i^{-1}\right)\right\} I(\boldsymbol{u} \in \bar{U}^*)
\end{aligned}
$$

$$\leqslant b \exp\left\{-\frac{1-2a}{2}\left(\sum_{i>j} q_{ij}^2 + \sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i\right)\right\} I(\boldsymbol{u} \in \bar{U}^*)$$
$$\leqslant b \exp\left(-\frac{1-2a}{2}\sum_{i>j} q_{ij}^2\right) \exp\left(-\frac{1-2a}{4}\sum_{i=1}^{p} f_i\right).$$

Consequently

$$|x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})| h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) \leqslant c_4 \prod_{j=1}^{p} f_j^{\frac{n-j-1}{2}} \exp\left(-\frac{1-2a}{2}\sum_{i>j} q_{ij}^2\right)$$
$$\times \exp\left(-\frac{1-2a}{4}\sum_{i=1}^{p} f_i\right).$$

Denote the right-hand side by $h(\boldsymbol{f}, \boldsymbol{q})$. Obviously

$$\int_{R_+^p} \int_{R^{\frac{p(p-1)}{2}}} h(\boldsymbol{f}, \boldsymbol{q})\, d\boldsymbol{q}\, d\boldsymbol{f} < \infty.$$

This guarantees the exchange between $\lim_{\rho\to 0}$ and the integral in (A.8); i.e.,

$$\lim_{\rho\to 0} I_1 = \int_{R_+^p} \int_{R^{\frac{p(p-1)}{2}}} \lim_{\rho\to 0} \{x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\, h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\}\, d\boldsymbol{q}\, d\boldsymbol{f}.$$

Notice that

$$\lim_{\rho\to 0} I\left(f_1\lambda_1 > \cdots > f_p\lambda_p\right) = 1, \quad \lim_{\rho\to 0} \prod_{j<i}\left(1 - \frac{f_i\lambda_i}{f_j\lambda_j}\right) = 1.$$

Since

$$\lim_{\rho\to 0} \boldsymbol{u}(\boldsymbol{\lambda}, \boldsymbol{f}, \boldsymbol{q}) = \boldsymbol{0}, \quad g_{ii}(\boldsymbol{0}) = 1,\ 1 \leqslant \forall i \leqslant p, \quad g_{ij}(\boldsymbol{0}) = 0,\ 1 \leqslant \forall i < \forall j \leqslant p,$$

we have

$$\lim_{\rho\to 0} I(\boldsymbol{u}(\boldsymbol{\lambda}, \boldsymbol{f}, \boldsymbol{q}) \in \bar{U}^*) = 1, \quad \lim_{\rho\to 0} J^*(\boldsymbol{u}(\boldsymbol{\lambda}, \boldsymbol{f}, \boldsymbol{q})) = J^*(\boldsymbol{0}),$$

and

$$\lim_{\rho\to 0} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{p} g_{ii}^2(\boldsymbol{u}) f_i + \sum_{i<j} g_{ij}^2(\boldsymbol{u}) f_j \lambda_j \lambda_i^{-1}\right)\right\} = \exp\left(-\frac{1}{2}\sum_{i=1}^{p} f_i\right).$$

Consequently,

$$\lim_{\rho\to 0} E[x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) \mid \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'] = \lim_{\rho\to 0} I_1$$
$$= c_3 J^*(\boldsymbol{0}) \int_{R_+^p} \int_{R^{\frac{p(p-1)}{2}}} \bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}) \prod_{i=1}^{p} f_i^{\frac{n-i-1}{2}}$$
$$\times \exp\left(-\frac{1}{2}\sum_{i>j} q_{ij}^2\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{p} f_i\right) d\boldsymbol{q}\, d\boldsymbol{f}. \tag{A.9}$$

Considering the special case $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) \equiv 1$, we notice that $c_3 J^*(\boldsymbol{0})$ is the normalizing constant for the joint distribution of $\boldsymbol{f}$ and $\boldsymbol{q}$, whose elements are all mutually independently distributed as $q_{ij} \sim N(0, 1),\ 1 \leqslant j < i \leqslant p,\ f_i \sim \chi^2_{n-i+1},\ 1 \leqslant i \leqslant p$. Therefore, the right side of (A.9) is equal to $E[\bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q})]$. □

Proof of Lemma 3. We write

$$\operatorname{tr} \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{\Sigma}^{-1} = \sum_{i,j} \widetilde{h}_{ij}^2 d_j/\lambda_i,$$

where $\widetilde{h}_{ij}$'s are the elements of $\widetilde{\boldsymbol{H}}(\boldsymbol{W}) = \boldsymbol{\Gamma}'\boldsymbol{H}(\boldsymbol{W})$. Since each term is non-negative, from Assumption 1, we have

$$\exists M,\ \forall \boldsymbol{\lambda}, \quad E(\widetilde{h}_{ij}^2 d_j/\lambda_i) < M.$$

By the Markov inequality

$$\forall c > 0,\ \forall \boldsymbol{\lambda}, \quad P(\widetilde{h}_{ij}^2 d_j/\lambda_i \geqslant c) \leqslant \frac{M}{c}.$$

Therefore,

$$\widetilde{h}_{ij}^2 d_j/\lambda_i = \widetilde{h}_{ij}^2 c_j (l_j/\lambda_j)(\lambda_j/\lambda_i) = O_p(1). \tag{A.10}$$

By Corollary 1 and Theorem 1, $r \xrightarrow{p} 0$ and $d(\boldsymbol{G}, \boldsymbol{\Gamma}) \xrightarrow{p} 0$. From these convergence and Assumption 1,

$$\frac{1}{c_j(\boldsymbol{G}, \boldsymbol{l})} \xrightarrow{p} \frac{1}{\bar{c}_j(\boldsymbol{\Gamma})},$$

which means

$$\frac{1}{c_j} = O_p(1). \tag{A.11}$$

From (A.10), (A.11) and $\lambda_j/l_j = O_p(1)$ (Lemma 1),

$$\widetilde{h}_{ij}^2 \xrightarrow{p} 0, \quad 1 \leqslant \forall j < \forall i \leqslant p,$$

i.e. $d(\boldsymbol{\Gamma}'\boldsymbol{H}(\boldsymbol{W}), \boldsymbol{I_p}) \xrightarrow{p} 0$. By the left orthogonal invariance of $d(\cdot,\cdot)$ in (4),

$$d(\boldsymbol{H}(\boldsymbol{W}), \boldsymbol{\Gamma}) \xrightarrow{p} 0. \qquad \square$$

AppendixB. Local coordinates of $\mathcal{O}(p)$ around $\boldsymbol{I_p}$

Here we discuss some details of local coordinates of $\mathcal{O}(p)$ around $\boldsymbol{I_p}$. We verify how the actual computation of our local coordinates can be carried out in principle, but general explicit formulas seem to be complicated.

First we verify the condition of implicit function theorem to show that $\boldsymbol{u}$ in (9) can be used as a local coordinate system around $\boldsymbol{I_p}$. Write

$$\begin{aligned}
\boldsymbol{\psi}_{ii}(\boldsymbol{G}) &= \sum_{t=1}^{p} g_{ti}^2 - 1, \quad 1 \leqslant i \leqslant p,\\
\boldsymbol{\psi}_{ij}(\boldsymbol{G}) &= \sum_{t=1}^{p} g_{ti} g_{tj}, \quad 1 \leqslant i < j \leqslant p,\\
\boldsymbol{\psi}(\boldsymbol{G}) &= (\boldsymbol{\psi}_{ij})_{1 \leqslant i \leqslant j \leqslant p}.
\end{aligned}$$

Then $\mathcal{O}(p)$ is defined by $\psi = \boldsymbol{0}$. Differentiate ψ with respect to $g_{11}, \ldots, g_{pp}$ and $g_{12}, \ldots, g_{p-1,p}$ and evaluate at $\boldsymbol{G} = \boldsymbol{I_p}$. Then we easily obtain

$$\left.\frac{\partial \psi(\boldsymbol{G})}{\partial(g_{11}, \ldots, g_{pp}, g_{12}, \ldots, g_{p-1,p})}\right|_{\boldsymbol{G}=\boldsymbol{I_p}} = \begin{pmatrix} 2\boldsymbol{I_p} & \boldsymbol{0} \\ \boldsymbol{0} & I_{\frac{p(p-1)}{2}} \end{pmatrix}.$$

The right-hand side is non-singular. The implicit function theorem tells us that there are two open sets $U,\ V$ and $\frac{p(p+1)}{2}$ C^∞ functions ϕ_{ij} on U $(1 \leqslant i \leqslant j \leqslant p)$ such that

$$\boldsymbol{0} \subset U \subset R^{\frac{p(p-1)}{2}}, \quad \boldsymbol{I_p} \in V \subset \mathcal{O}(p)$$

and the function $\boldsymbol{G}(\boldsymbol{u}) = (g_{ij}(\boldsymbol{u}))$ defined by

$$g_{ij}(\boldsymbol{u}) = \phi_{ij}(\boldsymbol{u}), \quad 1 \leqslant i \leqslant j \leqslant p, \quad g_{ij}(\boldsymbol{u}) = u_{ij}, \quad 1 \leqslant j < i \leqslant p,$$

is a one-to-one function from U onto V.

Next we consider Taylor expansion of $g_{ij} = \phi_{ij}(\boldsymbol{u}),\ i \leqslant j$, around the origin. It seems to be convenient to use the matrix exponential function. Let $\boldsymbol{Z} = (z_{ij})$ denote a skew symmetric matrix and let

$$\exp(\boldsymbol{Z}) = \boldsymbol{I_p} + \boldsymbol{Z} + \frac{1}{2!}\boldsymbol{Z}^2 + \cdots$$

be the matrix exponential function. Then $\boldsymbol{Z} \mapsto \exp(\boldsymbol{Z})$ defines a C^∞ diffeomorphism between a neighborhood of the origin in $R^{p(p-1)/2}$ and a neighborhood of $\boldsymbol{I_p}$ of $\mathcal{O}(p)$ (see [13, Section 9.5]). Consider the lower triangular part of

$$\boldsymbol{G} = \boldsymbol{I_p} + \boldsymbol{Z} + \frac{1}{2!}\boldsymbol{Z}^2 + \cdots.$$

Then for $i > j$

$$u_{ij} = z_{ij} + \frac{1}{2}\sum_k z_{ik}z_{kj} + \frac{1}{3!}\sum_{k,l} z_{ik}z_{kl}z_{lj} + \cdots. \tag{B.1}$$

For convenience define

$$u_{ii} = 0,\ 1 \leqslant i \leqslant p, \quad u_{ij} = -u_{ji},\ i < j.$$

Solving (B.1) by back substitution, we obtain

$$z_{ij} = -z_{ji} = u_{ij} - \frac{1}{2}\sum_k u_{ik}u_{kj} + \frac{1}{3}\sum_{k,l} u_{ik}u_{kl}u_{lj} + \cdots, \quad i > j.$$

From now on we only consider up to the second degree terms, because the third degree terms seem to be already somewhat cumbersome to handle. Then we have

$$g_{ii} = 1 - \frac{1}{2}\sum_j z_{ij}^2 + \cdots = 1 - \frac{1}{2}\sum_j u_{ij}^2 + \cdots,$$
$$g_{ij} = z_{ij} + \frac{1}{2}\sum_k z_{ik}z_{kj} + \cdots = u_{ij} + \sum_k u_{ik}u_{kj} + \cdots, \quad i < j.$$

Note that in Theorem 2, the orders of $u_{ij} = \tilde{g}_{ij}$ depend on the individual ratios λ_{i+1}/λ_i, $1 \leqslant i \leqslant p-1$. Therefore, it is difficult to simply express the asymptotic distribution of the upper triangular part of $\tilde{\boldsymbol{G}}$.

Finally, we discuss the differential form expression of the invariant measure on $\mathcal{O}(p)$ in terms of our local coordinates for small dimensions. For $p = 2$, write

$$\boldsymbol{G} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then $u = u_{21} = \sin\theta$ and

$$du = \cos\theta\, d\theta = \cos(\sin^{-1}(u))\, d\theta = \pm\sqrt{1-u^2}\, d\theta.$$

Furthermore

$$\boldsymbol{g}_1' d\boldsymbol{g}_2 = -(\cos^2\theta + \sin^2\theta)\, d\theta = -d\theta = \pm\frac{1}{\sqrt{1-u^2}}\, du.$$

Therefore $J^*(u) = 1/\sqrt{1-u^2}$. The case of $p = 2$ is obviously simple.

Now consider $p = 3$. Differentiating $\boldsymbol{G}'\boldsymbol{G} = \boldsymbol{I_p}$, we have $d\boldsymbol{G}'\boldsymbol{G} + \boldsymbol{G}'d\boldsymbol{G} = \boldsymbol{0}$, namely $\boldsymbol{G}'d\boldsymbol{G}$ is skew symmetric. Therefore

$$\bigwedge_{i>j} \boldsymbol{g}_j' d\boldsymbol{g}_i = \pm \bigwedge_{i<j} \boldsymbol{g}_j' d\boldsymbol{g}_i .$$

Now

$$\begin{aligned}(d\boldsymbol{g}_1, d\boldsymbol{g}_2) \\ = \begin{pmatrix} \frac{\partial g_{11}}{\partial u_{21}} du_{21} + \frac{\partial g_{11}}{\partial u_{31}} du_{31} + \frac{\partial g_{11}}{\partial u_{32}} du_{32} & \frac{\partial g_{12}}{\partial u_{21}} du_{21} + \frac{\partial g_{12}}{\partial u_{31}} du_{31} + \frac{\partial g_{12}}{\partial u_{32}} du_{32} \\ du_{21} & \frac{\partial g_{22}}{\partial u_{21}} du_{21} + \frac{\partial g_{22}}{\partial u_{31}} du_{31} + \frac{\partial g_{22}}{\partial u_{32}} du_{32} \\ du_{31} & du_{32} \end{pmatrix}\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{g}_2' d\boldsymbol{g}_1 &= g_{12} \left(\frac{\partial g_{11}}{\partial u_{21}} du_{21} + \frac{\partial g_{11}}{\partial u_{31}} du_{31} + \frac{\partial g_{11}}{\partial u_{32}} du_{32} \right) + g_{22} du_{21} + g_{32} du_{31} \\
&= \left(g_{12} \frac{\partial g_{11}}{\partial u_{21}} + g_{22} \right) du_{21} + \left(g_{12} \frac{\partial g_{11}}{\partial u_{31}} + u_{32} \right) du_{31} + g_{12} \frac{\partial g_{11}}{\partial u_{32}} du_{32} \\
&= \tau_{11}(\boldsymbol{u}) du_{21} + \tau_{12}(\boldsymbol{u}) du_{31} + \tau_{13}(\boldsymbol{u}) du_{32} \quad \text{(say)}, \\
\boldsymbol{g}_3' d\boldsymbol{g}_1 &= \left(g_{13} \frac{\partial g_{11}}{\partial u_{21}} + g_{23} \right) du_{21} + \left(g_{13} \frac{\partial g_{11}}{\partial u_{31}} + g_{33} \right) du_{31} + g_{13} \frac{\partial g_{11}}{\partial u_{32}} du_{32} \\
&= \tau_{21}(\boldsymbol{u}) du_{21} + \tau_{22}(\boldsymbol{u}) du_{31} + \tau_{23}(\boldsymbol{u}) du_{32} \quad \text{(say)}, \\
\boldsymbol{g}_3' d\boldsymbol{g}_2 &= \left(g_{13} \frac{\partial g_{12}}{\partial u_{21}} + g_{23} \frac{\partial g_{22}}{\partial u_{21}} \right) du_{21} + \left(g_{13} \frac{\partial g_{12}}{\partial u_{31}} + g_{23} \frac{\partial g_{22}}{\partial u_{31}} \right) du_{31} \\
&\quad + \left(g_{13} \frac{\partial g_{12}}{\partial u_{32}} + g_{23} \frac{\partial g_{22}}{\partial u_{32}} + g_{33} \right) du_{32} \\
&= \tau_{31}(\boldsymbol{u}) du_{21} + \tau_{32}(\boldsymbol{u}) du_{31} + \tau_{33}(\boldsymbol{u}) du_{32} \quad \text{(say)}.
\end{aligned}$$

Write $\tau(\boldsymbol{u}) = (\tau_{ij}(\boldsymbol{u}))_{1 \leqslant i,j \leqslant 3}$. Note that in $\boldsymbol{g}_2' d\boldsymbol{g}_1 \wedge \boldsymbol{g}_3' d\boldsymbol{g}_1 \wedge \boldsymbol{g}_3' d\boldsymbol{g}_2$ we only need to keep track of $du_{21} \wedge du_{31} \wedge du_{32}$. Then by the antisymmetry of the wedge product, we obtain the determinant of the 3×3 matrix $\tau(\boldsymbol{u})$ as the coefficient of $du_{21} \wedge du_{31} \wedge du_{32}$. Therefore

$$J^*(\boldsymbol{u}) = |\det \boldsymbol{\tau}(\boldsymbol{u})|.$$

Explicit expression of the right-hand side already seems to be complicated. On the other hand, it is clear that similar calculation can be carried out for a general dimension.

References

[1] T.W. Anderson, Asymptotic theory for principal component analysis, Ann. Math. Statist. 4 (1963) 122–148.

[2] O.B. Berger, Tail minimaxity in location vector problems and its applications, Ann. Statist. 4 (1976) 33–50.

[3] A.W. Davis, Asymptotic theory for principal component analysis: non-normal case, Austral. J. Statist. 19 (1977) 206–212.

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

[5] A.K. Gupta, V.L. Girko (Eds.), Multidimensional Statistical Analysis and Theory of Random Matrices: Proceedings of the Sixth Eugene Lukacs Symposium, VSP International Science Publishers, Zeist, The Netherlands, 1996.
[6] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematical and Statistical Problems, Vol. 1, University of California Press, Berkeley, CA, 1961, pp. 361–380.
[7] I.M. Johnstone, On the distribution of the largest eigenvalue in principal component analysis, Ann. Statist. 29 (2001) 295–327.
[8] D.A. Klain, G.-C. Rota, Introduction to Geometric Probability, Cambridge University Press, Cambridge, 1997.
[9] K. Krishnamoorthy, A.K. Gupta, Improved minimax estimation of a normal precision matrix, Canad. J. Statist. 17 (1989) 91–102.
[10] S. Kuriki, A. Takemura, Tail probabilities of the maxima of multilinear forms and their applications, Ann. Statist. 29 (2001) 328–371.
[11] A.W. Marshal, I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, San Diego, 1979.
[12] R.J. Muirhead, Latent roots and matrix variates: a review of some asymptotic results, Ann. Statist. 6 (1978) 5–33.
[13] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[14] R.J. Muirhead, Y. Chikuse, Asymptotic expansions for the joint and marginal distributions of the latent roots of the covariance matrix, Ann. Statist. 3 (1975) 1011–1017.
[15] I. Olkin, J.B. Selliah, Estimating covariances in a multivariate normal distribution, in: S.S. Gupta, D. Moore (Eds.), Statistical Decision Theory and Related Topics, Vol. 2, Academic Press, New York, 1977, pp. 313–326.
[16] N. Pal, Estimating the normal dispersion matrix and the precision matrix from a decision-theoretic point of view: a review, Statist. Papers 34 (1993) 1–26.
[17] D. Sharma, K. Krishnamoorthy, Orthogonal equivariant minimax estimators of bivariate normal covariance matrix and precision matrix, Calcutta Statist. Ass. Bull. 32 (1983) 23–45.
[18] Y. Sheena, On minimaxity of some orthogonally invariant estimators of bivariate normal dispersion matrix, J. Japan Statist. Soc. 32 (2002) 193–207.
[19] 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.
[20] M. Siotani, T. Hayakawa, Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Science Press, Columbus, OH, 1985.
[21] N. Sugiura, Derivatives of the characteristic root of a symmetric or a Hermitian matrix with two applications in multivariate analysis, Comm. Statist. 1 (1973) 393–417.
[22] N. Sugiura, H. Ishibayashi, Reference prior Bayes estimator for bivariate normal covariance matrix with risk comparison, Comm. Statist.—Theory Methods 26 (1997) 2203–2221.
[23] A. Takemura, Y. Sheena, Distribution of eigenvalues and eigenvectors of Wishart matrix when the population eigenvectors are infinitely dispersed, Mathematical Engineering Section Technical Report METR 2002-13, The University of Tokyo, 2002.
[24] D.E. Tyler, Asymptotic inference for eigenvectors, Ann. Statist. 9 (1981) 725–736.
[25] D.E. Tyler, The asymptotic distribution of principal component roots under local alternatives to multiple roots, Ann. Statist. 11 (1983) 1232–1242.
[26] C.M. Waternaux, Asymptotic distribution of the sample roots for a nonnormal population, Biometrika 63 (1976) 639–645.