Distribution of eigenvalues and eigenvectors of Wishart matrix when the population eigenvalues are infinitely dispersed

Akimichi Takemura
Department of Mathematical Informatics, University of Tokyo
and
Yo Sheena
Department of Mathematics and Statistics
Bowling Green State University
(On leave from Department of Economics, Shinshu University, Japan.)

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 minimax estimators is identical when the sample eigenvalues become infinitely dispersed.

Keywords and phrases
asymptotic distribution, covariance matrix, minimax estimator, quadratic loss, singular parameter, Stein's loss, tail minimaxity.

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 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 (1963) and developed by many authors. Higher order asymptotic expansions for the case of distinct population roots were given by Sugiura (1973) and Muirhead and Chikuse (1975). For a review of related works see Section 3 of Muirhead (1978) and Section 10.3 of Siotani et al. (1985). Large sample theory under non-normality was studied by Tyler (1981, 1983).

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 Kuriki and Takemura (2001) 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 Johnstone (2001) 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

$$\boldsymbol{W} = \boldsymbol{G}\boldsymbol{L}\boldsymbol{G}', \qquad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}', \tag{1}$$

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

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

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

$$a^k(c, ac, a^2c, \ldots, a^{p-1}c), \qquad a \downarrow 0,$$
$$(a^{j-1}c, a^{j-2}c, \ldots, ac, c, bc, \ldots, b^{p-j}c), \qquad a \uparrow \infty,\ b \downarrow 0,$$

where $k \in R$ and $c > 0$ are fixed. In the parameter space of positive definite covariance matrices, the case of infinitely dispersed population eigenvalues corresponds to an extreme boundary of the parameter space.

We investigate the asymptotic distribution of the sample eigenvalues $(l_1, \ldots, l_p)$ and the sample eigenvectors $\boldsymbol{G}$ under the the limiting process (2). We will prove that after appropriate standardization, $l_1, \ldots, l_p$ 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 main motivation of the above result is the investigation of the tail minimaxity in the sense of Berger (1976) 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 (1976) 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 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 minimax estimation of $\boldsymbol{\Sigma}$. All proofs of the results are given in Section 4. Some additional technical results are given in Appendix.

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, we prove the consistency of the sample eigenvectors in Theorem 1 and derive the asymptotic distribution in Theorem 2. Proofs of the results are given in Section 4.

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 $\epsilon > 0$ there exist $C_1, C_2,\ 0 < C_1 < C_2$, such that*

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

Note that C_1, C_2 above do note depend on $\boldsymbol{\Sigma}$. From this lemma 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*

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

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

$$r \xrightarrow{p} 0$$

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

$$\rho(\boldsymbol{\Sigma}) < \delta \quad \Rightarrow \quad P(r(\boldsymbol{W}) > \epsilon) < \epsilon.$$

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 $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\Gamma}$) 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 $\boldsymbol{\Sigma}$, it is cumbersome to state the consistency result for $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are identified if there exists a diagonal matrix $\boldsymbol{D} = \mathrm{diag}(\pm 1, \ldots, \pm 1)$ such that $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2 \boldsymbol{D}$. Write two arbitrary orthogonal matrices $\boldsymbol{\Gamma}_i,\ i = 1, 2$, as $\boldsymbol{\Gamma}_1 = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)$, $\boldsymbol{\Gamma}_2 = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$. The elements of $\mathcal{O}(p)/\{-1,1\}^p$ are sometimes called *frames* in the field of geometric probability (see Chapter 6 of Klain and Rota (1997)). The quotient topology of $\mathcal{O}(p)/\{-1,1\}^p$ induced from $\mathcal{O}(p)$ can be metrized, for example, by the squared distance

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

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

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

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 $\forall \epsilon > 0, \exists \delta > 0$,

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

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 $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$ have distinct roots. We assume that the signs of the columns of $\boldsymbol{\Gamma}$ are chosen in some way and fixed. Let

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

Following Anderson (1963), 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

$$\begin{aligned} f_i &= \frac{l_i}{\lambda_i}, \qquad 1 \le i \le p, \\ q_{ij} &= \widetilde{g}_{ij} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}} = \widetilde{g}_{ij} f_j^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}, \qquad 1 \le j < i \le p. \end{aligned}$$

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.

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

$$\begin{aligned} f_i &\xrightarrow{d} \chi^2_{n-i+1}, \qquad 1 \le i \le p, \\ q_{ij} &\xrightarrow{d} N(0,1), \qquad 1 \le j < i \le p, \end{aligned}$$

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

Concerning this theorem, we again discuss the indeterminacy of the signs of the eigenvectors. In this theorem we chose the signs of sample eigenvectors by requiring non-negativeness 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,

$$(g_{21},\ g_{31}, \ldots, g_{p1},\ g_{32}, \ldots, g_{p,p-1}) = \boldsymbol{u} = (u_{ij})_{1 \le j < i \le p}. \qquad (6)$$

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

$$x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = x(\boldsymbol{\Gamma}\boldsymbol{G}(\boldsymbol{u}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})), \boldsymbol{l}(\boldsymbol{f}, \boldsymbol{\lambda}), \boldsymbol{\lambda}), \tag{7}$$

where

$$\begin{aligned} &\boldsymbol{f} = (f_1, \ldots, f_p), \quad \boldsymbol{q} = (q_{ij})_{1 \le j < i \le p}, \quad \boldsymbol{u}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) = (q_{ij} f_j^{-\frac{1}{2}} \lambda_j^{-\frac{1}{2}} \lambda_i^{\frac{1}{2}})_{1 \le j < i \le p}, \\ &\boldsymbol{l}(\boldsymbol{f}, \boldsymbol{\lambda}) = (f_1 \lambda_1, \ldots, f_p \lambda_p). \end{aligned}$$

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

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

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*

$$\exists a < \frac{1}{2},\ \exists b > 0, \qquad \sup_{\boldsymbol{\lambda}} |x(\boldsymbol{\Gamma}\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda})| \le b \operatorname{etr}(a\boldsymbol{G}\boldsymbol{L}\boldsymbol{G}'\boldsymbol{\Lambda}^{-1}) \quad \textit{a.e. in } (\boldsymbol{G}, \boldsymbol{l}). \tag{8}$$

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. In this section we derive some necessary conditions for an estimator $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ to be minimax with respect to Stein's (entropy) loss function as well as the quadratic loss function. Assuming mild regularity conditions we prove one lemma and two theorems. Lemma 3 shows that estimated eigenvectors are consistent for estimators of bounded risks. Then in Theorem 3 and 4 we show that the behavior of a broad class of minimax estimators is identical when the sample eigenvalues become infinitely dispersed. This corresponds to a necessary condition for tail minimaxity in the sense of Berger (1976). For the case of the estimation of a location vector, Berger (1976) gave sufficient conditions of tail minimaxity in a general multivariate location family. Our result, being a necessary condition, is in a sense weaker than the result of Berger (1976). On the other hand, the problem of estimating $\boldsymbol{\Sigma}$ seems to be technically harder. It would be of interest to investigate sufficient conditions for tail minimaxity in our setting.

Before going into technical details, we prefer to present a motivation for our results on identical tail behavior of a broad class of minimax estimators. By definition, risks of minimax estimators have to approach the minimax risk somewhere in the parameter space. It seems reasonable to expect that risk functions of shrinkage type minimax estimators approach the minimax risk as $\rho(\boldsymbol{\Sigma}) \to 0$. Wishart distributions, being an exponential family, possess the completeness of the sufficient statistic $\boldsymbol{W}$, even if the parameter space is restricted to any open set, where ρ is small. The completeness suggests that minimax estimators with their risks close to the minimax risk should be close to each other. Clearly it might not be true that the risk function of every minimax estimator approaches the minimax risk smoothly as $\rho \to 0$. For example, consider multiple shrinkage type minimax estimators in the sense of George (1986a, b) with possibly infinite number of shrinkage points. The risk function of such an estimator might fluctuate even when $\rho \to 0$. In any case, the above argument is only a plausible motivation and below we need to impose several regularity conditions on existence of appropriate limits to justify our results.

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 Pal (1993). Stein's loss function is one of the most frequently used loss functions for the estimation of $\boldsymbol{\Sigma}$ and it is given by

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

The first minimax estimator was given by James and Stein (1961). It is defined by

$$\widehat{\boldsymbol{\Sigma}}^{JS} = \boldsymbol{T}\,\mathrm{diag}(\delta_1^{JS}, \ldots, \delta_p^{JS})\boldsymbol{T}', \tag{9}$$

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

$$\delta_i^{JS} = \frac{1}{n+p+1-2i}, \qquad 1 \le i \le p. \tag{10}$$

This type of estimators, $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{T}'$, $\boldsymbol{D} = \mathrm{diag}(\delta_1, \ldots, \delta_p)$, are called *triangularly equivariant*; i.e., for any lower triangular matrix $\boldsymbol{S}$ with positive diagonal elements, $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}') = \boldsymbol{S}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{S}'$. The estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ has the constant minimax risk, which is given by

$$\bar{R}_1 = -\sum_{i=1}^{p} \log \delta_i^{JS} - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}]. \tag{11}$$

Another important loss function is the quadratic loss function given by

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

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

$$a_{ij} = \begin{cases} (n+p-2i+1)(n+p-2i+3), & \text{if} \quad i = j, \\ (n+p-2i+1), & \text{if} \quad i > j, \\ (n+p-2j+1), & \text{if} \quad j > i, \end{cases} \tag{12}$$

$$b_i = n+p+1-2i, \quad i = 1, \ldots, p. \tag{13}$$

and let

$$\boldsymbol{\delta} = (\delta_1^{OS}, \ldots, \delta_p^{OS})' = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{14}$$

Then the minimax risk is given by

$$\bar{R}_2 = p - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} = p - \boldsymbol{\delta}'\boldsymbol{A}\boldsymbol{\delta}, \tag{15}$$

(Olkin and Selliah (1977), Sharma and Krishnamoorthy (1983)) and the estimator

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

has the constant minimax risk, where $\boldsymbol{T}$ is defined as in (9).

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

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

In accordance with the definition of $\boldsymbol{G}$, we determine the sign of $\boldsymbol{H}(\boldsymbol{W})$ by $(\boldsymbol{\Gamma}'\boldsymbol{H})_{ii} \geq 0,\ 1 \leq \forall i \leq p$. We also use the notation

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

An estimator of the form

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{G}\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{G}', \qquad \boldsymbol{\Psi}(\boldsymbol{L}) = \mathrm{diag}(\psi_1(\boldsymbol{l}), \ldots, \psi_p(\boldsymbol{l})). \tag{16}$$

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

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{W}) &= \boldsymbol{G}, \\ c_i(\boldsymbol{W}) &= c_i(\boldsymbol{l}) = \frac{\psi_i(\boldsymbol{l})}{l_i}, \qquad 1 \leq i \leq p. \end{aligned}$$

Here we mention the orthogonally equivariant minimax estimator $\widehat{\boldsymbol{\Sigma}}^{SDS}$ derived independently by Stein and by Dey and Srinivasan (1985). This estimator is defined by

$$\psi_i(\boldsymbol{l}) = l_i\delta_i^{JS}, \qquad 1 \leq i \leq p,$$

where δ_i^{JS} is given in (10). $\widehat{\boldsymbol{\Sigma}}^{SDS}$ is of simple form but has substantially better risk than the M.L.E., $\boldsymbol{W}/n$, and also dominates $\widehat{\boldsymbol{\Sigma}}^{JS}$ with respect to Stein's loss. See Dey and Srinivasan (1985) and Sugiura and Ishibayashi (1997) for more details. Order preservation among $\psi_i(\boldsymbol{l}),\ i = 1, \ldots, p$, is discussed in Sheena and Takemura (1992).

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

$$\psi_i(\boldsymbol{l}) = l_i \delta_i^{OS}, \qquad 1 \le i \le p$$

with δ_i^{OS} in (14) has been considered to be minimax from an analogy between $\widehat{\boldsymbol{\Sigma}}^{SDS}$ and $\widehat{\boldsymbol{\Sigma}}^{KG}$. (See Krishnamoorthy and Gupta (1989).) For the case $p = 2$, this conjecture was proved by Sheena (2001).

Now we introduce some regularity conditions on $\widehat{\boldsymbol{\Sigma}}$. 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 $\rho \to 0$, $\boldsymbol{G}$ converges to $\boldsymbol{\Gamma}$ in probability and hence the risk function of an estimator should depend only on $\boldsymbol{G}$ close to $\boldsymbol{\Gamma}$.

First we make an assumption on the convergence of $c_i(\boldsymbol{G}, \boldsymbol{l})$.

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

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

Next we assume boundedness of the following expectation.

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

$$E[\mathrm{tr}\, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{\Sigma}^{-1}] < M \tag{17}$$

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

Under these assumptions we have the following result.

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 Assumptions 1 and 2. Then as $\rho = \rho(\boldsymbol{\Lambda}) \to 0$,*

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

Note that $\mathrm{tr}\, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})\boldsymbol{\Sigma}^{-1}$ 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 (17). Note that a minimax estimator has a bounded risk by definition.

Note that $\boldsymbol{G} \xrightarrow{p} \boldsymbol{\Gamma}$ by Theorem 1 and $\boldsymbol{H}(\boldsymbol{W}) \xrightarrow{p} \boldsymbol{\Gamma}$ by Lemma 3. Therefore $\boldsymbol{H}(\boldsymbol{W}) - \boldsymbol{G} \xrightarrow{p} 0$. Since $r \xrightarrow{p} 0$ by Corollary 1, this intuitively implies $\boldsymbol{H}(\boldsymbol{W}) = \boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}) \to \boldsymbol{\Gamma}$ as $\boldsymbol{G} \to \boldsymbol{\Gamma}$ and $r \to 0$. However at present it seems difficult to prove this implication without assuming the existence of the limit. Therefore we state another assumption concerning the existence of the limit of $\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l})$.

Assumption 3 *There exists* $\bar{\boldsymbol{H}}(\boldsymbol{\Gamma}) \in \mathcal{O}(p)$*, such that*

$$\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}) \to \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})$$

as $\boldsymbol{G} \to \boldsymbol{\Gamma}$ *and* $r \to 0$.

Combining Assumption 3 and Lemma 3, we can easily prove the following theorem.

Theorem 3 *Let* $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W})$ *be an estimator satisfying Assumptions 1, 2 and 3. Then* $\bar{\boldsymbol{H}}(\boldsymbol{\Gamma}) = \boldsymbol{\Gamma}$.

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., $\widetilde{h}_{ij}(\boldsymbol{G}, \boldsymbol{l}) \to 0$, $1 \le \forall j < \forall i \le p$. 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

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

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 $1 \le j < i \le p$. 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

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

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

Assumption 4 *There exist* $\bar{\zeta}_{ij,\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q})$, $1 \le j < i \le p$, *such that*

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

Note that this assumption implies $\widetilde{h}_{ij,\boldsymbol{\Gamma}} \to 0$ as $\rho \to 0$ for $1 \le \forall j < \forall i \le p$.

The following lemma gives the asymptotic lower bounds of the risks, $R_i(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = E[L_i(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$, $i = 1, 2$.

Lemma 4 *Under Assumptions 1 and 4*

$$\varliminf_{\rho \to 0} R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Gamma\Lambda\Gamma'}) \ \ge\ \sum_{i=1}^{p}\sum_{j=1}^{p} \bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ij}^2] - \sum_{i=1}^{p} \log \bar{c}_i(\boldsymbol{\Gamma}) - \sum_{i=1}^{p} E[\log \chi^2_{n-i+1}] - p, \tag{18}$$

$$\varliminf_{\rho \to 0} R_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Gamma\Lambda\Gamma'}) \ \ge\ \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \bar{c}_i(\boldsymbol{\Gamma})\bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ki}\xi_{kj}\xi_{li}\xi_{lj}] - 2\sum_{i=1}^{p}\sum_{j=1}^{p} \bar{c}_j(\boldsymbol{\Gamma}) E[\xi_{ij}^2] + p, \tag{19}$$

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

$$\exists a < \frac{1}{8},\ \exists b > 0,\ 1 \le \forall i, \forall j \le p, \quad \zeta_{ij}(\boldsymbol{\Gamma G}, \boldsymbol{l}) \le b \operatorname{etr}(a\boldsymbol{GLG'\Lambda}^{-1}) \qquad a.e.\ in\ (\boldsymbol{G}, \boldsymbol{l}), \tag{20}$$

and

$$0 < \exists c_l < \exists c_u < \infty,\ 1 \le \forall i \le p, \quad c_l < c_i(\boldsymbol{W}) < c_u, \tag{21}$$

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* (18) $(i = 1)$ *or* (19) $(i = 2)$*.*

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

$$c_l \boldsymbol{W} \le \widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}) \le c_u \boldsymbol{W}, \tag{22}$$

then $c_l < c_i(\boldsymbol{W}) < c_u,\ i = 1 \le \forall i \le p$, by the minimax characterization of the eigenvalues (see for example Marshal and Olkin (1979) Ch.20 A.1.b). In particular (22) holds for triangularly equivariant estimators $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{T} \operatorname{diag}(\delta_1, \ldots, \delta_p)\boldsymbol{T}'$ with $c_l = \min(\delta_1, \ldots, \delta_p)$ and $c_u = \max(\delta_1, \ldots, \delta_p)$.

Lemma 4 shows that the values of $\bar{\zeta}_{ij,\boldsymbol{\Gamma}},\ 1 \le j < i \le p$, determine the risk or its lower bound. We now make the final assumption.

Assumption 5 *In Assumption 4*

$$\bar{\zeta}_{ij,\boldsymbol{\Gamma}} = 1, \quad 1 \le \forall j < \forall i \le p. \tag{23}$$

Note that any orthogonally equivariant estimator satisfies this condition, since $\widetilde{h}_{ij} = \widetilde{g}_{ij}$ and

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

Therefore Assumption 5 seems to imply that we are restricted to estimators which are nearly orthogonally equivariant. However as we argue below, Assumption 5 should hold for a broad class of estimators. As a particular case, in Appendix A we confirm that 2×2 triangularly equivariant estimators satisfy Assumption 5 except for the following particular $\boldsymbol{\Gamma}$

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{pmatrix}.$$

Except for this $\boldsymbol{\Gamma}$, a 2×2 triangularly equivariant estimator behaves like an orthogonally equivariant estimator as $r \to 0$. The singularity at this particular $\boldsymbol{\Gamma}$ seems to be an interesting fact and also suggests the difficulty of relaxing our regularity conditions. We conjecture that triangularly equivariant estimators for general dimension p satisfies Assumption 5 for almost all $\boldsymbol{\Gamma}$, although at present we do not have a proof.

We now argue why $\bar{\zeta}_{ij,\boldsymbol{\Gamma}} = 1$ should hold for many estimators. Consider $\widetilde{\boldsymbol{H}}(\boldsymbol{u}, \boldsymbol{l}) = (\widetilde{h}_{ij}(\boldsymbol{u}, \boldsymbol{l})) = \widetilde{\boldsymbol{H}}(\boldsymbol{\Gamma G}(\boldsymbol{u}), \boldsymbol{l})$, the function of $(\boldsymbol{u}, \boldsymbol{l})$. If Assumptions 1–3 hold for the neighborhood $\boldsymbol{\Gamma G}(U)$ of $\boldsymbol{\Gamma}$, then by Theorem 3 $\widetilde{\boldsymbol{H}}(\boldsymbol{u}, \boldsymbol{l}) \to \boldsymbol{G}(\boldsymbol{u})$ on U as $r \to 0$; i.e., $\widetilde{h}_{ij}(\boldsymbol{u}, \boldsymbol{l}) \to u_{ij}$ on U for $1 \le j < i \le p$.

For sufficiently smooth estimators, the derivatives (up to a certain order) of $\widetilde{h}_{ij}(\boldsymbol{u}, \boldsymbol{l})$ with respect to $\boldsymbol{u}$ at $\boldsymbol{u} = \boldsymbol{0}$ converge to the corresponding derivatives of u_{ij}; i.e., for $1 \le j_l < i_l \le p,\ l = 1, \ldots, k$,

$$\lim_{r \to 0} \frac{\partial^k \widetilde{h}_{ij}}{\partial u_{i_1 j_1} \cdots \partial u_{i_k j_k}}(\boldsymbol{0}, \boldsymbol{l}) = \frac{\partial^k u_{ij}}{\partial u_{i_1 j_1} \cdots \partial u_{i_k j_k}}(\boldsymbol{0}) = \begin{cases} 1, & \text{if } \ k = 1,\ u_{i_1 j_1} = u_{ij}, \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

In the case of Stein's loss it can be shown that slow convergence in (24) incurs risk penalty. This implies that for nice estimators, among the terms of Taylor expansion of $\widetilde{h}_{ij}$ with respect to $\boldsymbol{u}$ around $\boldsymbol{0}$, those with the order of $\lambda_i^{\frac{1}{2}}/\lambda_j^{\frac{1}{2}}$ after the substitution of $\boldsymbol{u} = (q_{kl} f_l^{-\frac{1}{2}} \lambda_l^{-\frac{1}{2}} \lambda_k^{\frac{1}{2}})_{1 \le l < k \le p}$ determine $\zeta_{ij,\boldsymbol{\Gamma}}$. Although many terms might have this order (e.g. $u_{is} u_{sj},\ j < \exists s < i$), by (24) the only term of which coefficient does not vanish as $r \to 0$ is u_{ij}. Since its coefficient converges to 1, $\zeta_{ij,\boldsymbol{\Gamma}} \to 1$.

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

Theorem 4 *Suppose that $\widehat{\boldsymbol{\Sigma}}$ satisfies Assumptions 1 and 5. If it is minimax with respect to Stein's loss, then*

$$c_i^*(\boldsymbol{\Gamma}) = \delta_i^{JS}, \qquad i = 1, \ldots, p,$$

where δ_i^{JS} is given in (10)*. If it is minimax with respect to the quadratic loss, then*

$$c_i^*(\boldsymbol{\Gamma}) = \delta_i^{OS}, \qquad i = 1, \ldots, p,$$

where δ_i^{OS} is given in (14)*.*

Roughly speaking, Theorem 3 and 4 indicate that when r is very small, any minimax estimator satisfying Assumptions 1–5 must be approximately same as $\widehat{\boldsymbol{\Sigma}}^{SDS}$ in the case of Stein's loss and $\widehat{\boldsymbol{\Sigma}}^{KG}$ in the case of the quadratic loss respectively; i.e.,

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

We state one simple application of Theorem 4. The orthogonally equivariant estimators given by (16) contain the subclass (say C^o) of estimators which is defined by

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

This class contains the M.L.E., $\boldsymbol{W}/n$, and $\widehat{\boldsymbol{\Sigma}}^{SDS}$. We have conjectured that $\widehat{\boldsymbol{\Sigma}}^{SDS}$ and $\widehat{\boldsymbol{\Sigma}}^{OS}$ are the only minimax estimators in this class with respect to Stein's loss and the quadratic loss. This conjecture was proved in Sheena (2001) for the case $p = 2$. It is obvious that every estimator in C^o satisfies Assumptions 1 and 5. Then from Theorem 4, it is necessary that

$$c_i = \delta_i^{JS(OS)}, \qquad 1 \le i \le p,$$

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

4 Proofs

In this section we present proofs of the results of the previous sections.

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

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

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

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

we have

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

Hence

$$P\left(\frac{l_1}{\lambda_1} \ge C_2\right) \le P\left(\chi^2_{np} \ge C_2\right), \quad \forall C_2 > 0. \tag{25}$$

This means

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

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 $\boldsymbol{\Lambda}_{22} = \operatorname{diag}(\lambda_2, \ldots, \lambda_p)$. Let $\tilde{l}_2$ be the largest eigenvalue of $\boldsymbol{W}_{22}$. Then by the minimax characterization of eigenvalues (see for example Marshal and Olkin (1979) Ch.20 A.1.c)

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

If we use the result (25) for $\tilde{l}_2/\lambda_2$, we have

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

Therefore

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

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

$$P\left(\frac{l_i}{\lambda_i} \ge C_2\right) \to 0 \quad \text{as } C_2 \to \infty, \quad 1 \le \forall i \le p, \tag{26}$$

uniformly in $\boldsymbol{\Lambda}$.

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

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

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

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

we have

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

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

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

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

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

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

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

We also have

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

Therefore

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

and

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

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

$$P\left(\frac{l_i}{\lambda_i} \leq C_1\right) \to 0 \quad \text{as } \ C_1 \to 0, \ \ 1 \leq \forall i \leq p, \tag{27}$$

uniformly in $\boldsymbol{\Lambda}$.

From (26),(27) and Bonferroni inequality, we can choose C_1 and C_2 for any given $\epsilon > 0$ so that

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

∎

Proof of Theorem 1

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

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

Therefore

$$\tilde{g}_{ij}^2 \le \widetilde{w}_{ii}\frac{\lambda_i}{l_j} = \widetilde{w}_{ii}\frac{\lambda_j}{l_j}\frac{\lambda_i}{\lambda_j} \le \widetilde{w}_{ii}\frac{\lambda_j}{l_j}\rho. \tag{28}$$

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

$$P(\widetilde{w}_{ii} < M) > 1 - \epsilon, \qquad \forall \boldsymbol{\Sigma}. \tag{29}$$

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

$$P\left(\left|\frac{\lambda_j}{l_j}\right| < C\right) > 1 - \epsilon, \qquad \forall \boldsymbol{\Sigma}. \tag{30}$$

From (29) and (30), we can easily prove

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

From this fact and (28) we have

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

∎

Proof of Lemma 2

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

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

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

$$
\begin{aligned}
&E[x(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})] \\
&\quad = E[x(\boldsymbol{\Gamma}\widetilde{\boldsymbol{G}},\boldsymbol{l},\boldsymbol{\lambda})] \\
&\quad = c_1 \prod_{i=1}^{p} \lambda_i^{-\frac{n}{2}} \int_{\mathcal{L}} \int_{\mathcal{O}(p)^+} x(\boldsymbol{\Gamma G},\boldsymbol{l},\boldsymbol{\lambda}) \prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} \prod_{j<i} (l_j - l_i) \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{GLG'\Lambda}^{-1}\right) d\mu(\boldsymbol{G}) d\boldsymbol{l}, \quad (31)
\end{aligned}
$$

where

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

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 $g_{ii}(\boldsymbol{0}) = 1,\ 1 \le \forall i \le p$, there exists an open set U^* that satisfies

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

and

$$g_{ii}(\boldsymbol{u}) \ge \sqrt{\frac{1}{2}}, \quad 1 \le \forall i \le p, \quad \forall \boldsymbol{u} \in \bar{U}^*. \quad (32)$$

Let

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

The integral (31) is divided into two parts, say I_1 over $\mathcal{N}(\boldsymbol{I_p})$ and I_2 over $\mathcal{N}(\boldsymbol{I_p})^C$. Then from (8)

$$
\begin{aligned}
I_2 &\le c_1 \prod_{i=1}^{p} \lambda_i^{-\frac{n}{2}} \int_{\mathcal{N}(I_p)^C} |x(\boldsymbol{\Gamma G},\boldsymbol{l},\boldsymbol{\lambda})| \prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} \prod_{j<i} (l_j - l_i) \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{GLG'\Lambda}^{-1}\right) d\mu(\boldsymbol{G}) d\boldsymbol{l} \\
&\le c_1\, b \prod_{i=1}^{p} \lambda_i^{-\frac{n}{2}} \int_{\mathcal{N}(I_p)^C} \prod_{i=1}^{p} l_i^{\frac{n-p-1}{2}} \prod_{j<i} (l_j - l_i)\, \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{GLG'}\widetilde{\boldsymbol{\Lambda}}^{-1}\right) d\mu(\boldsymbol{G}) d\boldsymbol{l} \\
&= c_1'\, P\left(\boldsymbol{\Gamma' G} \notin \mathcal{N}(\boldsymbol{I_p}) \mid \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma'}\right), \quad (33)
\end{aligned}
$$

where $\widetilde{\boldsymbol{\Lambda}} = (1-2a)\boldsymbol{\Lambda}$. Since $P\left(\boldsymbol{\Gamma' G} \notin \mathcal{N}(\boldsymbol{I_p}) \mid \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma'}\right) \to 0$ as $\rho \to 0$ by Theorem 1, I_2 vanishes.

Now we focus ourselves on I_1.

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

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

$$c_2 \bigwedge_{i>j} \boldsymbol{g}_j' d\boldsymbol{g}_i, \quad (34)$$

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

$$\begin{aligned} dg_{ij} &= du_{ij}, \quad i > j, \\ dg_{ij} &= \sum_{k>l} \frac{\partial g_{ij}}{\partial u_{kl}} du_{kl}, \quad i \le j, \end{aligned}$$

into (34) 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

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

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

$$\begin{aligned} \boldsymbol{f} &= (f_i)_{1 \le i \le p}, \qquad f_i = \frac{l_i}{\lambda_i}, \\ \boldsymbol{q} &= (q_{ij})_{1 \le j < i \le p}, \qquad q_{ij} = u_{ij} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}. \end{aligned}$$

The Jacobian of this transformation is

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

Furthermore

$$\begin{aligned} \prod_{i=1}^p l_i^{\frac{n-p-1}{2}} &= \prod_{i=1}^p f_i^{\frac{n-p-1}{2}} \prod_{i=1}^p \lambda_i^{\frac{n-p-1}{2}}, \\ \prod_{j<i} (l_j - l_i) &= \prod_{j<i} (f_j \lambda_j - f_i \lambda_i) = \prod_{j<i} f_j \lambda_j \left(1 - \frac{f_i \lambda_i}{f_j \lambda_j} \right) \end{aligned}$$

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

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

$$
I_1 = \int_{R^{p(p-1)/2}} \int_{R_+^p} x_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda}) h(\boldsymbol{f}, \boldsymbol{q}, \boldsymbol{\lambda})\, d\boldsymbol{f} d\boldsymbol{q}, \tag{35}
$$

where

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

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 $J^*(\boldsymbol{u})\, I(\boldsymbol{u}) \le K$ for some $K > 0$, because $J^*(\boldsymbol{u})$ is bounded on a compact set $\bar{U}^*$ as remarked above. Clearly

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

From (8) we have

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

Therefore, using (32), we have

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

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

Consequently

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

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 (35); 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

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

Since

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

we have

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

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

$$
\begin{aligned}
&\lim_{\rho \to 0} E[x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) \mid \boldsymbol{\Sigma} = \boldsymbol{\Gamma\Lambda\Gamma'}] \\
&\quad = \ \lim_{\rho \to 0} I_1 \\
&\quad = \ c_3 \ J^*(\boldsymbol{0}) \int_{R_+^p} \int_{R^{\frac{p(p-1)}{2}}} \bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q}) \prod_{i=1}^{p} f_i^{\frac{n-i-1}{2}} \exp\left(-\frac{1}{2}\sum_{i>j} q_{ij}^2\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{p} f_i\right) d\boldsymbol{q} d\boldsymbol{f}. \qquad (36)
\end{aligned}
$$

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 $q_{ij} \sim N(0, 1), \ 1 \le j < i \le p, \ f_i \sim \chi^2_{n-i+1}, \ 1 \le i \le p$. Therefore the right side of (36) is equal to $E[\bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{f}, \boldsymbol{q})]$. ∎

Proof of Theorem 2

Let A denote the event

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

with arbitrary real numbers $\beta_{ij}, \; 1 \le j \le i \le p$. Define $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) = I(A)$, where $I(\cdot)$ is the indicator function. Then from the definition of $x_{\boldsymbol{\Gamma}}$ in (7)

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

Note that $x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) = I(A)$ satisfies the inequality (8) with $a = 0, \; b = 1$. Therefore by the Lemma 2, we have the result for any $\boldsymbol{\Gamma}$. ∎

Proof of Lemma 3

We write

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

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,

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

By Markov inequality

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

Therefore

$$\widetilde{h}_{ij}^2 d_j / \lambda_i = \widetilde{h}_{ij}^2 c_j (l_j / \lambda_j)(\lambda_j / \lambda_i) = O_p(1). \tag{37}$$

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

$$\frac{1}{c_j} = O_p(1). \tag{38}$$

From (37), (38) and $\lambda_j / l_j = O_p(1)$ (Lemma 1),

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

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),

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

∎

Proof of Theorem 3

Consider the triangular inequality

$$d(\boldsymbol{\Gamma}, \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})) \leq d(\boldsymbol{\Gamma}, \boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l})) + d(\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}), \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})). \tag{39}$$

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 $d(\boldsymbol{G}, \boldsymbol{\Gamma}) \xrightarrow{p} 0$ as $\rho \to 0$. Therefore by Assumption 3, $d(\boldsymbol{H}(\boldsymbol{G}, \boldsymbol{l}), \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})) \xrightarrow{p} 0$. It follows that the right hand side of (39) converges to 0 in probability as $\rho \to 0$. However the left hand side is a constant and therefore $d(\boldsymbol{\Gamma}, \bar{\boldsymbol{H}}(\boldsymbol{\Gamma})) = 0$. ∎

Proof of Lemma 4

Here for notational simplicity we prove the lemma 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 $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}'\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Gamma}\boldsymbol{W}\boldsymbol{\Gamma}')\boldsymbol{\Gamma}$ 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 $\alpha_{ij} = \alpha_{ij}(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}) = h_{ij}(\boldsymbol{G}, \boldsymbol{l})c_j^{\frac{1}{2}} l_j^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}}$, $1 \leq i, j \leq p$, we can write the loss functions as

$$\begin{aligned} L_1(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}, \boldsymbol{\Lambda}) &= \operatorname{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}') - \sum_{i=1}^{p} \log c_i - \log|\boldsymbol{\Lambda}^{-1}\boldsymbol{W}| - p \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{ij}^2 - \sum_{i=1}^{p} \log c_i - \log|\boldsymbol{\Lambda}^{-1}\boldsymbol{W}| - p. \end{aligned}$$

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

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

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

By Fatou's lemma, we have

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

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 4 and the fact $h_{ii}\to 1,\ 1\le\forall i\le p,\ h_{ij}\to 0,\ 1\le\forall i<\forall j\le p$, we have

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

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 (20) and (21). Since $|\alpha_{ij}(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})| \le c_u^{\frac{1}{2}}|\zeta_{ij}(\boldsymbol{G},\boldsymbol{l})||g_{ij}|l_j^{\frac{1}{2}}\lambda_i^{-\frac{1}{2}}$,

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

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

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

$$|\alpha_{ki}(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})\alpha_{kj}(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})\alpha_{li}(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})\alpha_{lj}(\boldsymbol{G},\boldsymbol{l},\boldsymbol{\lambda})| \le (bb')^4\mathrm{etr}\{4(a+a')\boldsymbol{GLG'\Lambda}^{-1}\}\ a.e.,$$

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

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

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

∎

Proof of Theorem 4 for Stein's loss

Substitute

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

into (18). Then we have

$$\begin{aligned} &\varliminf_{\rho \to 0} R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) \\ &\geq \sum_{i>j} \bar{c}_j(\boldsymbol{\Gamma}) E[\chi_1^2] + \sum_{i=1}^p \bar{c}_i(\boldsymbol{\Gamma}) E[\chi_{n-i+1}^2] - \sum_{i=1}^p \log \bar{c}_i(\boldsymbol{\Gamma}) - \sum_{i=1}^p E\left[\log \chi_{n-i+1}^2\right] - p \\ &= \sum_{i>j} \bar{c}_j(\boldsymbol{\Gamma}) + \sum_{i=1}^p \bar{c}_i(\boldsymbol{\Gamma})(n-i+1) - \sum_{i=1}^p \log \bar{c}_i(\boldsymbol{\Gamma}) - \sum_{i=1}^p E\left[\log \chi_{n-i+1}^2\right] - p \\ &= \sum_{i=1}^p \left\{ \bar{c}_j(\boldsymbol{\Gamma})/\delta_i^{JS} - \log \bar{c}_j(\boldsymbol{\Gamma}) \right\} - \sum_{i=1}^p E\left[\log \chi_{n-i+1}^2\right] - p \end{aligned}$$

with δ_i^{JS} given in (10). The right side is uniquely minimized when $\bar{c}_i(\boldsymbol{\Gamma}) = \delta_i^{JS}$. Therefore

$$\varliminf_{\rho \to 0} R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) \geq \sum_{i=1}^p \log(n+p+1-2i) - \sum_{i=1}^p E\left[\log \chi_{n-i+1}^2\right],$$

where the right side is equal to the minimax risk $\bar{R}_1$ in (11). If $\bar{c}_i(\boldsymbol{\Gamma}) \neq \delta_i, \ 1 \leq \exists i \leq p$, then there exists some $\boldsymbol{\Sigma}$ such that

$$R_1(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) > \bar{R}_1.$$

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

$$c_i^*(\boldsymbol{\Gamma}) = \delta_i^{JS}, \qquad 1 \leq \forall i \leq p.$$

∎

Proof of Theorem 4 for the quadratic loss

The proof is straightforward but long. We briefly sketch the outline. We decompose the four-folded summation in (19) as

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

Furthermore

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

We also have

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

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

$$
\varliminf_{\rho\to 0} R_2(\widehat{\boldsymbol{\Sigma}},\boldsymbol{\Gamma\Lambda\Gamma'}) \geq \boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} - 2\boldsymbol{b}'\boldsymbol{c} + p, \tag{40}
$$

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 (12) and (13). The minimum of the quadratic function (40) is uniquely attained when $\boldsymbol{c}$ satisfies the linear equation $\boldsymbol{Ac} = \boldsymbol{b}$, namely

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

where δ_i^{OS} is defined in (14) and the minimum value of the quadratic function is the minimax risk $\bar{R}_2$ in (15). This proves the theorem for the case of the quadratic loss. ∎

A The case of 2×2 triangularly equivariant estimator

Here we confirm that 2×2 triangularly equivariant estimators satisfy our regularity conditions of Section 3, except for the case that the (1,1)-element of $\boldsymbol{\Gamma}$ is 0.

Let

$$
\boldsymbol{W}_\Gamma = \boldsymbol{\Gamma G}(u)\boldsymbol{L G'}(u)\boldsymbol{\Gamma'},
$$

where

$$\boldsymbol{G}(u) = \begin{pmatrix} \sqrt{1-u^2} & -u \\ u & \sqrt{1-u^2} \end{pmatrix}, \qquad u = q_{21} f_1^{-\frac{1}{2}} \lambda_2^{\frac{1}{2}} \lambda_1^{-\frac{1}{2}}.$$

Write $y = (\lambda_2/\lambda_1)^{\frac{1}{2}}$. If the triangularly equivariant estimator based on $\boldsymbol{W}_\Gamma$ and its spectral decomposition is given by

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}_\Gamma) = \boldsymbol{T}\,\mathrm{diag}(\delta_1, \delta_2)\boldsymbol{T}' = \boldsymbol{HDH'}$$

with

$$\boldsymbol{W}_\Gamma = \boldsymbol{TT}',$$

then

$$\widehat{\boldsymbol{\Sigma}}(l_1^{-1}\boldsymbol{W}_\Gamma) = \boldsymbol{T}\,\mathrm{diag}(l_1^{-1}\delta_1, l_1^{-1}\delta_2)\boldsymbol{T}' = \boldsymbol{H}\widetilde{\boldsymbol{D}}\boldsymbol{H'},$$

where

$$\widetilde{\boldsymbol{D}} = \mathrm{diag}(\widetilde{d}_1, \widetilde{d}_2) = \mathrm{diag}(l_1^{-1}d_1, l_1^{-1}d_2).$$

Let $\widetilde{w}_{ij} = (l_1^{-1}\boldsymbol{W}_\Gamma)_{ij}$. Note that $\widetilde{h}_{21}$, $\widetilde{d}_1$ are C^∞ functions of $\widetilde{w}_{11}, \widetilde{w}_{12}, \widetilde{w}_{22}$ except the point $\widetilde{w}_{11} = 0$. Let γ_{11} denote the $(1,1)$-element of $\boldsymbol{\Gamma}$. We have to consider the cases $\gamma_{11} \neq 0$ and $\gamma_{11} = 0$ separately.

First consider the case $\gamma_{11} \neq 0$. We would like to show as $y \to 0$

$$\widetilde{h}_{21}/u \to 1, \qquad \widetilde{d}_1 \to \delta_1. \tag{41}$$

If $\gamma_{11} \neq 0$, we can eliminate the singularity at $\widetilde{w}_{11} = 0$ for sufficiently small y. In this case $\widetilde{h}_{21}$, $\widetilde{d}_1$ are C^∞ functions of $\widetilde{w}_{11}, \widetilde{w}_{12}, \widetilde{w}_{22}$. Since $u = O(y)$, for the calculation of (41) we can substitute $l_1^{-1}\boldsymbol{W}_\Gamma$ with any C^∞ matrix-valued function of y, $\boldsymbol{W}_\Gamma^*$ such that $\boldsymbol{W}_\Gamma^* = \boldsymbol{W}_\Gamma + o(y)$. Since

$$\begin{aligned}
\boldsymbol{G}(u) l_1^{-1}\boldsymbol{LG'}(u) &= \begin{pmatrix} \sqrt{1-u^2} & -u \\ u & \sqrt{1-u^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & f_2 f_1^{-1} y^2 \end{pmatrix} \begin{pmatrix} \sqrt{1-u^2} & u \\ -u & \sqrt{1-u^2} \end{pmatrix} \\
&= \begin{pmatrix} 1 - u^2 + f_2 f_1^{-1} y^2 u^2 & u\sqrt{1-u^2}(1 - f_2 f_1^{-1} y^2) \\ u\sqrt{1-u^2}(1 - f_2 f_1^{-1} y^2) & u^2 + (1-u^2) f_2 f_1^{-1} y^2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & u \\ u & u^2 \end{pmatrix} + o(y),
\end{aligned}$$

we can substitute $l_1^{-1}\boldsymbol{W}_\Gamma$ with

$$\boldsymbol{W}_\Gamma^* = \boldsymbol{\Gamma} \begin{pmatrix} 1 & u \\ u & u^2 \end{pmatrix} \boldsymbol{\Gamma'}.$$

Then

$$\boldsymbol{W}_\Gamma^* = \boldsymbol{TT}', \qquad \boldsymbol{T} = \boldsymbol{\Gamma} \begin{pmatrix} 1 & 0 \\ u & 0 \end{pmatrix}$$

is the lower-triangular decomposition of $\boldsymbol{W}_\Gamma^*$ and

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}(\boldsymbol{W}_\Gamma^*) &= \boldsymbol{\Gamma} \begin{pmatrix} 1 & 0 \\ u & 0 \end{pmatrix} \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix} \begin{pmatrix} 1 & u \\ 0 & 0 \end{pmatrix} \boldsymbol{\Gamma'} \\
&= \boldsymbol{\Gamma} O(u) \begin{pmatrix} (1+u^2)\delta_1 & 0 \\ 0 & 0 \end{pmatrix} O'(u) \boldsymbol{\Gamma'},
\end{aligned} \tag{42}$$

where $O(u)$ is the orthogonal matrix given by

$$O(u) = \begin{pmatrix} \dfrac{1}{\sqrt{1+u^2}} & \dfrac{-u}{\sqrt{1+u^2}} \\ \dfrac{u}{\sqrt{1+u^2}} & \dfrac{1}{\sqrt{1+u^2}} \end{pmatrix}.$$

The right side of (42) is the spectral decomposition of $\widehat{\mathbf{\Sigma}}(\boldsymbol{W}_\Gamma^*)$. Therefore

$$\boldsymbol{H} = \mathbf{\Gamma} O(u), \qquad \widetilde{\boldsymbol{D}} = \mathrm{diag}(\sqrt{1+u^2}\delta_1, 0)$$

Hence

$$\widetilde{\boldsymbol{H}} = \mathbf{\Gamma}'\boldsymbol{H} = O(u).$$

Since

$$\frac{\widetilde{h}_{21}}{u} = \frac{1}{\sqrt{1+u^2}} \to 1$$

and

$$\widetilde{d}_1 = \sqrt{1+u^2}\delta_1 \to \delta_1,$$

(41) is proved.

Now we will prove $c_2 \to \delta_2$. Note that

$$d_1 d_2 = |\widehat{\mathbf{\Sigma}}(\boldsymbol{W}_\Gamma)| = \delta_1\delta_2|\boldsymbol{T}\boldsymbol{T}'| = \delta_1\delta_2 l_1 l_2,$$

$$c_1 c_2 = \frac{d_1 d_2}{l_1 l_2} = \delta_1\delta_2.$$

Since $c_1 \to \delta_1$,

$$c_2 \to \delta_2.$$

Next, we consider the case $\gamma_{11} = 0$, i.e.

$$\mathbf{\Gamma} = \begin{pmatrix} 0 & \epsilon_1 \\ \epsilon_2 & 0 \end{pmatrix} \qquad \epsilon_i = \pm 1.$$

In this case the eigenvectors and eigenvalues exhibit certain singularity. In fact we show that the relevant limits depend on how $\boldsymbol{G}$ converges to $\mathbf{\Gamma}$. Since

$$l_1^{-1}\boldsymbol{W}_\Gamma = \begin{pmatrix} u^2 + (1-u^2)f_2 f_1^{-1}y^2 & \epsilon_1\epsilon_2 u\sqrt{1-u^2}(1-f_2 f_1^{-1}y^2) \\ \epsilon_1\epsilon_2 u\sqrt{1-u^2}(1-f_2 f_1^{-1}y^2) & 1-u^2+f_2 f_1^{-1}y^2u^2 \end{pmatrix},$$

$\widehat{\mathbf{\Sigma}}(l_1^{-1}\boldsymbol{W}_\Gamma) = (\hat{\sigma}_{ij})$ is given by

$$\begin{aligned} \hat{\sigma}_{11} &= \delta_1\{u^2 + (1-u^2)f_2 f_1^{-1}y^2\}, \\ \hat{\sigma}_{12} &= \delta_1\epsilon_1\epsilon_2 u\sqrt{1-u^2}(1-f_2 f_1^{-1}y^2), \\ \hat{\sigma}_{22} &= (\delta_1-\delta_2)\frac{u^2(1-u^2)(1-f_2 f_1^{-1}y^2)^2}{u^2+(1-u^2)f_2 f_1^{-1}y^2} + \delta_2\{(1-u^2)+f_2 f_1^{-1}y^2u^2\}. \end{aligned} \tag{43}$$

Note that

$$\hat{\sigma}_{11} \to 0,$$
$$\hat{\sigma}_{12} \to 0,$$
$$\hat{\sigma}_{22} \to \delta_1 \frac{q_{21}^2}{f_2 + q_{21}^2} + \delta_2 \frac{f_2}{f_2 + q_{21}^2}.$$

Therefore

$$\begin{aligned} \widetilde{d}_1 + \widetilde{d}_2 &= \hat{\sigma}_{11} + \hat{\sigma}_{22} \to \delta_1 \frac{q_{21}^2}{f_2 + q_{21}^2} + \delta_2 \frac{f_2}{f_2 + q_{21}^2}, \quad (44) \\ \widetilde{d}_1 \widetilde{d}_2 &= \hat{\sigma}_{11}\hat{\sigma}_{22} - \hat{\sigma}_{12}^2 \to 0. \quad (45) \end{aligned}$$

From (44) and (45),

$$c_1 = \widetilde{d}_1 \to \delta_1 \frac{q_{21}^2}{f_2 + q_{21}^2} + \delta_2 \frac{f_2}{f_2 + q_{21}^2}.$$

Since $c_1 c_2 = \delta_1 \delta_2$,

$$c_2 \to \frac{\delta_1 \delta_2 (f_2 + q_{21}^2)}{\delta_1 q_{21}^2 + \delta_2 f_2}.$$

Now we consider the convergence of $\widetilde{h}_{21}$. From $(\boldsymbol{\Gamma}'\boldsymbol{H})_{ii} \geq 0$, $\boldsymbol{H}$ is of the form

$$\boldsymbol{H} = \begin{pmatrix} \epsilon_1 \tau & \epsilon_1 \sqrt{1-\tau^2} \\ \epsilon_2 \sqrt{1-\tau^2} & -\epsilon_2 \tau \end{pmatrix}.$$

Notice $\widetilde{h}_{21} = \tau$ and

$$\begin{aligned} \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{21} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} \end{pmatrix} &= \begin{pmatrix} \epsilon_1 \tau & \epsilon_1 \sqrt{1-\tau^2} \\ \epsilon_2 \sqrt{1-\tau^2} & -\epsilon_2 \tau \end{pmatrix} \begin{pmatrix} \widetilde{d}_1 & 0 \\ 0 & \widetilde{d}_2 \end{pmatrix} \begin{pmatrix} \epsilon_1 \tau & \epsilon_2 \sqrt{1-\tau^2} \\ \epsilon_1 \sqrt{1-\tau^2} & -\epsilon_2 \tau \end{pmatrix} \\ &= \begin{pmatrix} \tau^2(\widetilde{d}_1 - \widetilde{d}_2) + \widetilde{d}_2 & \epsilon_1\epsilon_2 \tau\sqrt{1-\tau^2}(\widetilde{d}_1 - \widetilde{d}_2) \\ \epsilon_1\epsilon_2 \tau\sqrt{1-\tau^2}(\widetilde{d}_1 - \widetilde{d}_2) & \tau^2(\widetilde{d}_2 - \widetilde{d}_1) + \widetilde{d}_1 \end{pmatrix}. \quad (46) \end{aligned}$$

From (43) and (46),

$$\epsilon_1\epsilon_2 \tau\sqrt{1-\tau^2}(\widetilde{d}_1 - \widetilde{d}_2) = \delta_1 \epsilon_1 \epsilon_2 u \sqrt{1-u^2}(1 - f_2 f_1^{-1} y^2).$$

From this equation, it is obvious that $\tau > 0$ for sufficiently small y. Therefore $\widetilde{h}_{21}/u \to \sqrt{\alpha}$ is equivalent to $\widetilde{h}_{21}^2/u^2 = \tau^2/u^2 \to \alpha$. From the diagonal elements of (46),

$$\tau^2 = \frac{1}{2}\left(1 + \frac{\hat{\sigma}_{11} - \hat{\sigma}_{22}}{\widetilde{d}_1 - \widetilde{d}_2}\right) = \frac{1}{2}\left(1 + \frac{\hat{\sigma}_{11} - \hat{\sigma}_{22}}{\sqrt{(\hat{\sigma}_{11} + \hat{\sigma}_{22})^2 - 4\hat{\sigma}_{11}\hat{\sigma}_{22} + 4\hat{\sigma}_{12}^2}}\right).$$

If we consider sufficiently small $z = y^2$, $\hat{\sigma}_{11}, \hat{\sigma}_{22}, \hat{\sigma}_{12}^2$ are C^∞ functions of z, and τ^2 is a C^∞ function of $\hat{\sigma}_{11}, \hat{\sigma}_{22}, \hat{\sigma}_{12}^2$. Hence τ^2 is a C^∞ function of z in a neighborhood of $z = 0$. Put $g(z) = \tau^2(z)$. Then

$$\lim_{y \to 0} \frac{\tau^2}{y^2} = \lim_{z \to 0} \frac{g(z)}{z} = g'(0).$$

The calculation of $g'(z)$ is very long. With assistance of a computer software, we have

$$g'(0) = \frac{(f_2 + q_{21}^2)^2 \delta_1^2 q_{21}^2}{f_1(\delta_1 q_{21}^2 + \delta_2 f_2)^2}.$$

Therefore

$$\widetilde{h}_{21}/u \to \frac{\delta_1(q_{21}^2 + f_2)}{\delta_1 q_{21}^2 + \delta_2 f_2}.$$

B Local coordinates of orthogonal group around the identity matrix

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 (6) can be used as a local coordinate system around $\boldsymbol{I_p}$. Write

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

Then $\mathcal{O}(p)$ is defined by $\boldsymbol{\psi} = \boldsymbol{0}$. Differentiate $\boldsymbol{\psi}$ 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

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

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 $(1 \le i \le j \le p)$ such that

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

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

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

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

Next we consider Taylor expansion of $g_{ij} = \phi_{ij}(\boldsymbol{u})$, $i \le j$, 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

$$\exp(\boldsymbol{Z}) = \boldsymbol{I_p} + \boldsymbol{Z} + \frac{1}{2!}\boldsymbol{Z}^2 + \cdots \tag{47}$$

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 Section 9.5 of Muirhead (1982)). Consider the lower triangular part of

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

Then for $i > j$

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

For convenience define

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

Solving (47) 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

$$\begin{aligned} g_{ii} &= 1 - \tfrac{1}{2}\textstyle\sum_j z_{ij}^2 + \cdots = 1 - \tfrac{1}{2}\sum_j u_{ij}^2 + \cdots, \\ g_{ij} &= z_{ij} + \tfrac{1}{2}\textstyle\sum_k z_{ik}z_{kj} + \cdots = u_{ij} + \sum_k u_{ik}u_{kj} + \cdots, \qquad i < j. \end{aligned}$$

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

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

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

and

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

Write $\boldsymbol{\tau}(\boldsymbol{u}) = (\tau_{ij}(\boldsymbol{u}))_{1 \le i,j \le 3}$. 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 $\boldsymbol{\tau}(\boldsymbol{u})$ 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.

Acknowledgements
The second author is deeply grateful to the Department of Mathematics and Statistics, Bowling Green State University for providing comfortable research facilities.

References

[1] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, **36**, 413–432.

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

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

[4] George, E. I. (1986a). Minimax multiple shrinkage estimation, *Ann. Statist.*, **14**, 188–205.

[5] George, E. I. (1986b). Combining minimax shrinkage estimators. *Journal of the American Statistical Association*, **81**, 437–445.

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

[7] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.*, **29**, 295–327.

[8] Klain, D. A. and Rota, G-C. (1997). *Introduction to Geometric Probability.* Cambridge University Press, Cambridge.

[9] Krishnamoorthy, K. and Gupta, A. K. (1989). Improved minimax estimation of a normal precision matrix. *Canad. J. Statist.*, **17**, 91–102.

[10] Kuriki, S. and Takemura, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications. *Ann. Statist.*, **29**, 328–371.

[11] Marshal, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, San Diego.

[12] Muirhead, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results. *Ann. Statist.*, **6**, 5–33.

[13] Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* John Wiley & Sons, New York.

[14] Muirhead, R. J. and Chikuse, Y. (1975). Asymptotic expansions for the joint and marginal distributions of the latent roots of the covariance matrix. *Ann. Statist.*, **3**, 1011–1017.

[15] Olkin, I. and Selliah, J. B. (1977). Estimating covariances in a multivariate normal distribution. *Statistical Decision Theory and Related Topics*, **2**, 313-326, ed. Gupta, S. S. and Moore, D., Academic Press, New York.

[16] Pal, N. (1993). Estimating the normal dispersion matrix and the precision matrix from a decision-theoretic point of view: a review. *Statist. Papers*, **34**, 1-26.

[17] Sharma, D. and Krishnamoorthy, K. (1983). Orthogonal equivariant minimax estimators of bivariate normal covariance matrix and precision matrix. *Calcutta Statist. Ass. Bull.*, **32**, 23–45.

[18] Sheena, Y. (2001). On minimaxity of some orthogonally invariant estimators of bivariate normal dispersion matrix. *Department of Mathematics and Statistics Technical Report 01-13.* Bowling Green State University.

[19] Sheena, Y. and Takemura, A. (1992). Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss. *Journal of Multivariate Analysis*, **41**, 117–131.

[20] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook.* American Science Press, Columbus, Ohio.

[21] Sugiura, N. (1973). Derivatives of the characteristic root of a symmetric or a Hermitian matrix with two applications in multivariate analysis. *Comm. Statist.*, **1**, 393–417.

[22] Sugiura, N. and Ishibayashi, H. (1997). Reference prior Bayes estimator for bivariate normal covariance matrix with risk comparison. *Comm. Statist.–Theory Meth.*, **26**, 2203-2221.

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

[24] Tyler, D. E. (1983). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.*, **11**, 1232–1242.