Admissibility and minimaxity of Bayes estimators for a normal mean matrix

Hisayuki Tsukuma

Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan

Received 19 June 2007
Available online 13 February 2008

Abstract

In some invariant estimation problems under a group, the Bayes estimator against an invariant prior has equivariance as well. This is useful notably for evaluating the frequentist risk of the Bayes estimator. This paper addresses the problem of estimating a matrix of means in normal distributions relative to quadratic loss. It is shown that a matricial shrinkage Bayes estimator against an orthogonally invariant hierarchical prior is admissible and minimax by means of equivariance. The analytical improvement upon every over-shrinkage equivariant estimator is also considered and this paper justifies the corresponding positive-part estimator preserving the order of the sample singular values.

AMS 2000 subject classifications: primary 62C10; secondary 62C15; 62C20; 62J07

Keywords: Admissibility; Bayes estimation; Inadmissibility; Isotonic regression; Minimaxity; Order statistic; Quadratic loss; Simultaneous estimation; Singular value decomposition; Shrinkage estimator

1. Introduction

For $i = 1, \ldots, p$ and $j = 1, \ldots, m$, let $x_{ij}$ be random variable distributed as the normal distribution with unknown mean $\theta_{ij}$ and variance one. Suppose $p \leq m$ and $x_{ij}$’s are mutually independent. The simultaneous estimation of $\theta_{ij}$’s is then considered under the sum of the quadratic loss functions, $\sum_{i=1}^{p} \sum_{j=1}^{m} (\delta_{ij} - \theta_{ij})^2$, where $\delta_{ij}$’s are, respectively, certain estimators of $\theta_{ij}$’s. Let us write the above model as

$$X \sim \mathcal{N}_{p \times m}(\Theta, I_p \otimes I_m).$$

(1.1)
Here \( X = (x_{ij}) \), \( \Theta = (\theta_{ij}) \) and \( I_k \) is the identity matrix of order \( k \). The notation \( N_{p \times m}(\Theta, I_p \otimes I_m) \) denotes the matrix-variate normal distribution with mean matrix \( \Theta \) and covariance matrix \( I_p \otimes I_m \), where \( \otimes \) means the Kronecker product. Hence in other words, it is of great interest in this paper to estimate the mean matrix \( \Theta \) under the quadratic loss function (the squared Frobenius norm)

\[
\| \delta - \Theta \|^2 = \text{tr} \,(\delta - \Theta)(\delta - \Theta)^t, \tag{1.2}
\]

where \( \delta = (\delta_{ij}) \) and also \( \text{tr} A \) and \( A^t \) indicate the trace and the transpose of a square matrix \( A \), respectively.

Let \( \mathcal{O}_p \) be the group of \( p \)-dimensionally orthogonal matrices and let \( \mathcal{V}_{m,p} \) be the Stiefel manifold, namely, \( \mathcal{V}_{m,p} = \{ V \in R^{m \times p} \mid V^t V = I_p \} \). Write the singular value decomposition of \( X \) as \( ULV \), where \( U \in \mathcal{O}_p \), \( V \in \mathcal{V}_{m,p} \) and \( L = \text{diag}(l_1, \ldots, l_p) \) with \( l_1 > \cdots > l_p > 0 \). The estimation problem given above is invariant with respect to the group of transformations \( X \rightarrow PXQ \) and \( \Theta \rightarrow P\Theta Q \) for any \( P \in \mathcal{O}_p \) and \( Q \in \mathcal{O}_m \) and then the equivariant estimator of \( \Theta \) can be represented as

\[
\delta = \delta(X) = U \Psi(L)V^t, \tag{1.3}
\]

where \( \Psi(L) \) is a diagonal matrix whose \( i \)-th diagonal element \( \psi_i(L) \) is a function of \( L \). Hereafter we denote \( \Psi(L) = \text{diag} (\psi_1(L), \ldots, \psi_p(L)) \).

In the estimation problem of the mean matrix \( \Theta \), natural and usual estimator of \( \Theta \) is \( X \). In fact, \( X \) is the unbiased and maximum likelihood estimator and it is minimax with constant risk \( mp \). However, \( X \) has a weak point that it is improved on by certain shrinkage-type equivariant estimators

\[
\delta^s = X - (U \Phi(L^2)U^t)X = U \Psi^s(L)V^t, \quad \Psi^s(L) = L(I_p - \Phi(L^2)), \tag{1.4}
\]

where \( \Psi^s(L) = \text{diag} (\psi_1^s(L), \ldots, \psi_p^s(L)) \) and also \( \Phi(L^2) = \text{diag} (\phi_1(L^2), \ldots, \phi_p(L^2)) \). Etron and Morris [5] considered empirical Bayes estimation and proposed the shrinkage estimator with matricial shrinkage factor \(-(m - p - 1)(XX^t)^{-1}\),

\[
\delta^{EM} = (I_p - (m - p - 1)(XX^t)^{-1})X = UL(I_p - \Phi^{EM})V^t = U \Psi^{EM}V^t,
\]

where \( \Psi^{EM} = \text{diag} (\psi_1^{EM}, \ldots, \psi_p^{EM}) \) and \( \Phi^{EM} = \text{diag} (\phi_1^{EM}, \ldots, \phi_p^{EM}) \) with \( \phi_i^{EM} = (m - p - 1)/l_i^2 \). \( \delta^{EM} \) has smaller risk than \( X \) and is thus minimax. Subsequently, [13] developed the unbiased risk estimate formula and used it to pursue certain alternative estimators. [13] stated that \( \delta^{EM} \) is dominated by

\[
\delta^{ST} = UL(I_p - \Phi^{ST})V^t = U \Psi^{ST}V^t,
\]

where \( \Psi^{ST} = \text{diag} (\psi_1^{ST}, \ldots, \psi_p^{ST}) \) and \( \Phi^{ST} = \text{diag} (\phi_1^{ST}, \ldots, \phi_p^{ST}) \) with \( \phi_i^{ST} = d_i/l_i^2 \) and \( d_i = m + p - 2i - 1 \).

The estimators \( \delta^{EM} \) and \( \delta^{ST} \) are minimax, but there are two problems. First, the diagonal elements of \( \Psi^{EM} \) and \( \Psi^{ST} \) may often take negative values and second, the \( \psi_i^{ST} \)'s may not keep the order such as \( \psi_1^{ST} \geq \cdots \geq \psi_p^{ST} \). These seem unnatural situations since the singular values have both ordering and nonnegativity properties \( l_1 > \cdots > l_p > 0 \). Throughout this paper, if \( \psi_i^s(L) \geq \cdots \geq \psi_p^s(L) \) with probability one then \( \delta^s \) is called order-preserving and if \( \psi_i^s(L) < 0 \) or \( \phi_i(L^2) > 1 \) for some \( i \) then \( \delta^s \) is said to be overshrinking. In the event that some elements of \( \Psi^s(L) \) are negative, define the positive-part of \( \psi_i^s(L) \) as \( \psi_i^+(L) = \max(0, \psi_i^s(L)) \)
and denote $\Psi_+^s(L) = \text{diag}(\psi_1^+(L), \ldots, \psi_p^+(L))$. An estimator replacing $\Psi(L)$ by $\Psi_+^s(L)$ is named positive-part estimator.

There are some studies associated with positive-part estimators and, for instance, [5] showed that $\delta^{EM}$ is dominated by its positive-part estimator. For examples of more complicated estimators, see [17,18]). It is not known, however, whether a nonorder-preserving equivariant estimator is dominated by the corresponding one modifying the order and also whether there exists an alternative estimator blessed with both order-preserving and nonnegativity properties. The main objectives of this paper are to answer these questions and, moreover, to find out a superior estimator from the decision-theoretical point of view.

The Bayesian method is a powerful technique in decision theory; in particular, hierarchical models have been investigated recently. In hierarchical Bayes estimation of the normal mean vector, plenty of minimax and admissible estimators are obtained by Berger and Robert [1], Berger and Strawderman [2], Kubokawa and Strawderman [8]. In the case of the mean matrix $\Theta$, Berger et al. [3] led to general results on admissible hierarchical Bayes estimators, but did not treat the minimaxity of the Bayes estimators.

Section 2 of this paper handles admissible and minimax estimation of $\Theta$ with order-preserving and nonnegativity properties based on hierarchical Bayes procedure. To derive admissible and minimax estimators, we consider orthogonally invariant priors $\pi(\Theta)$ satisfying $\pi(\Theta) = \pi(P\Theta Q)$ for every $P \in O_p$ and $Q \in O_m$, and first show that the resulting Bayes estimators belong to the class of the equivariant estimators (1.3) without the disagreeable points like nonorder-preserving and negativity properties. These crucial facts are applied to prove the admissibility and minimaxity of a matricial shrinkage Bayes estimator against proper/improper hierarchical prior, which is natural multivariate extension of Strawderman’s [14]’s prior and different from Berger et al.’s [3] priors.

In Section 3 we establish improvement methods upon any nonorder-preserving and overshrinkage estimator such as $\delta^{EM}$ and $\delta^{ST}$. The similar argument as in [12] is used in order to show that, under the quadratic loss function (1.2), order-preserving and positive-part estimators are better than their original nonorder-preserving and overshrinkage estimators. Based on these results and the unbiased risk estimator due to Stein [13], some new minimax estimators are given. We also discuss the inadmissibility of a generalized Bayes estimator against an improper hierarchical prior considered in Section 2.

2. Bayes procedure for admissible and minimax estimation

2.1. Characterization of the Bayes estimators against orthogonally invariant priors

Let $\pi(\Theta)$ be a density function of prior distribution of $\Theta$. Assume here that $\pi(\Theta)$ is orthogonally invariant, namely,

$$\pi(\Theta) = \pi(P\Theta Q)$$  \hspace{1cm} (2.1)

for every $P \in O_p$ and $Q \in O_m$. Then the (generalized) Bayes estimator with respect to the quadratic loss (1.2) is equivalent to the posterior mean,

$$\delta^\pi = \delta^\pi(X) = \frac{\int \Theta \exp(- (1/2) \| \Theta - X \|^2) \pi(\Theta)(d\Theta)}{\int \exp(- (1/2) \| \Theta - X \|^2) \pi(\Theta)(d\Theta)}$$

with $(d\Theta) = \prod_{i=1}^p \prod_{j=1}^m d\theta_{ij}$. 
In some invariant estimation problem under a group, the Bayes estimator against an invariant prior has equivariance (see [9]). It is indeed easy to check that \( \delta^\pi(PXQ) = P\delta^\pi(X)Q \) for any \( P \in \mathcal{O}_p \) and \( Q \in \mathcal{O}_m \), so that the Bayes estimator \( \delta^\pi \) with the orthogonally invariant prior (2.1) is equivariant and has the form (1.3).

Furthermore, we are now able to identify the characteristic of the Bayes estimator \( \delta^\pi \).

**Lemma 2.1.** If \( \pi(\Theta) \) is an orthogonally invariant prior satisfying (2.1), then the resulting Bayes estimator \( \delta^\pi \) can be represented as

\[
\delta^\pi = U\Psi^\pi(L)V^\pi,
\]

where \( \Psi^\pi(L) = \text{diag}(\psi_1^\pi(L), \ldots, \psi_p^\pi(L)) \) with \( \psi_1^\pi(L) \geq \cdots \geq \psi_p^\pi(L) \geq 0 \). That is, \( \delta^\pi \) is an equivariant estimator with both order-preserving and nonnegativity properties.

**Proof.** Since \( \delta^\pi \) is equivariant, we will show only that the elements of \( \Psi^\pi(L) \) are order-preserving and nonnegative.

Let \( V_* \) be an \( m \times m \) orthogonal matrix such that \( V^\pi V_* = [I_p, 0_{p \times (m-p)}] \) and put \([\Psi^\pi(L), 0_{p \times (m-p)}] = U^\pi \delta^\pi V_* \). Using equivariance of \( \delta^\pi = \delta^\pi(X) \) yields

\[
U^\pi \delta^\pi(X) V_* = \delta^\pi(U^\pi XV_*) = \delta^\pi(L_*),
\]

where \( L_* = [L, 0_{p \times (m-p)}] \). Hence in order to prove that \( \psi_1^\pi(L) \geq \cdots \geq \psi_p^\pi(L) \geq 0 \), we must be able to ensure that for \( 1 \leq i < j \leq p \)

\[
\int (\theta_{ii} - \theta_{jj}) \exp \left( -\frac{1}{2} \| \Theta \|^2 + \text{tr} \Theta L^1_{*} \right) \pi(\Theta)(d\Theta) \geq 0 \tag{2.2}
\]

and that

\[
\int \theta_{pp} \exp \left( -\frac{1}{2} \| \Theta \|^2 + \text{tr} \Theta L^1_{*} \right) \pi(\Theta)(d\Theta) \geq 0, \tag{2.3}
\]

where \( \Theta = (\theta_{ij}) \).

To prove (2.2), let \( O_{ij} \) be the \( p \times p \) permutation matrix which interchanges the \( i \)-th and the \( j \)-th rows by the transformation \( \Theta \to O_{ij} \Theta \) and, analogously, let \( O_{ij}^* \) be the \( m \times m \) permutation matrix. Making the transformation \( \Theta \to O_{ij} \Theta O_{ij}^* \), we can write (2.2) as

\[
\int (\theta_{jj} - \theta_{ii}) \exp \left( l_i \theta_{jj} + l_j \theta_{ii} + \sum_{k \neq i,j} l_k \theta_{kk} - \frac{1}{2} \| \Theta \|^2 \right) \pi(\Theta)(d\Theta) \geq 0.
\]

Adding (2.2) to the above inequality gives that

\[
\int (\theta_{ii} - \theta_{jj})(e^{l_i \theta_{ii} + l_j \theta_{jj}} - e^{l_i \theta_{jj} + l_j \theta_{ii}}) \exp \left( \sum_{k \neq i,j} l_k \theta_{kk} - \frac{1}{2} \| \Theta \|^2 \right) \pi(\Theta)(d\Theta) \geq 0. \tag{2.4}
\]

Therefore we will show the above inequality.

Noting that \( l_i \theta_{ii} + l_j \theta_{jj} - (l_i \theta_{jj} + l_j \theta_{ii}) = (l_i - l_j)(\theta_{ii} - \theta_{jj}) \) and \( l_1 > \cdots > l_p > 0 \), we see that for \( i < j \)

\[
\begin{cases}
e^{l_i \theta_{ii} + l_j \theta_{jj}} \geq e^{l_i \theta_{jj} + l_j \theta_{ii}} & \text{if } \theta_{ii} - \theta_{jj} \geq 0, \\ e^{l_i \theta_{ii} + l_j \theta_{jj}} < e^{l_i \theta_{jj} + l_j \theta_{ii}} & \text{if } \theta_{ii} - \theta_{jj} < 0.
\end{cases}
\]
Hence the left-hand side of (2.4) is obviously nonnegative, which implies that $\Psi^\pi(L)$ is order-preserving.

To prove (2.3), let $P_p = \text{diag}(1, \ldots, 1, -1)$ and consider the transformation $Q \rightarrow P_pQ$. This result is combined with (2.3), so it suffices only to show that

$$
\int \theta_{pp}(e^{\theta_{pp}} - e^{-I_p\theta_{pp}}) \exp \left( \sum_{k=1}^{p-1} l_k \theta_{kk} - \frac{1}{2} \| \theta \|^2 \right) \pi(\theta)(d\theta) \geq 0. 
$$

(2.5)

It is here observed that

$$
e^{\theta_{pp}} \geq e^{-I_p\theta_{pp}} \quad \text{if } \theta_{pp} \geq 0,
$$

$$
e^{\theta_{pp}} < e^{-I_p\theta_{pp}} \quad \text{if } \theta_{pp} < 0,
$$

which results in the left-hand side of (2.5) being nonnegative, that is, the diagonal elements of $\Psi^\pi(L)$ being nonnegative. Thus the proof is complete. ■

Lemma 2.1 claims that the (generalized) Bayes estimators against orthogonally invariant priors lie in the class of equivariant estimators (1.3) and necessarily meet both order-preserving and nonnegativity properties. Now, let us consider the shrinkage-type estimators (1.4), that is, $\delta^s = U \Psi^s(L)V^t$, where $\Psi^s(L) = \text{diag}(\psi_1^s(L), \ldots, \psi_p^s(L))$. It follows from [4] that if $\delta^s$ is admissible then $\delta^s$ is a Bayes estimator. Thus, Lemma 2.1 suggests that if there exists an admissible Bayes estimator against an orthogonally invariant prior, of the form $\delta^s$, then it invariably satisfies $\psi_1^s(L) \geq \cdots \geq \psi_p^s(L) \geq 0$ with probability one.

In the case of $p = 1$ in the model (1.1), namely, in estimation of a mean vector of the multivariate normal distribution, the [7] estimator

$$
\delta^{JS} = X \left( 1 - \frac{m-2}{XX^t} \right) \equiv X\psi^{JS}(XX^t)
$$

is called the shrinkage estimator and dominates the maximum likelihood estimator $X$ relative to the quadratic loss. Because $\text{Pr}(\psi^{JS}(XX^t) < 0) = 0$, there is some possibility of $\delta^{JS}$ over-shrinking toward the zero vector. This is an unacceptable fact, and $\delta^{JS}$ is actually improved upon by its positive-part estimator $X \max(0, \psi^{JS}(XX^t))$ (see [9]). On the other hand Strawderman and Cohen [15] showed that the generalized Bayes estimator $\delta^\pi$ against an orthogonally invariant prior, $\pi(\theta) = \pi(\theta|Q)$ for $Q \in \mathcal{O}_m$, is rewritten as $\delta^\pi = X\psi(XX^t)$, where $\psi(\cdot)$ is a real-valued function. As in Lemma 2.1, it is easy to show that $\psi(\cdot) \geq 0$. Thus the orthogonally invariant Bayes estimator does not change the sign of each element of $X$, in other words it is always nonovershrinking.

2.2. Admissible and minimax estimators

Consider a hierarchical prior $\pi(\theta) = \int \pi(\theta|A)\pi(A)(dA)$ where $A = (\lambda_{ij})$ is a $p \times p$ positive definite matrix of hyperparameters and $(dA) = \bigwedge_{i \leq j} d\lambda_{ij}$. Suppose here $\theta|A \sim N_{p \times m}(\theta_{0p \times m}, A^{-1}(I_p - A) \otimes I_m)$, that is, the density of $\theta$ given $A$, $\pi(\theta|A)$, is specified by

$$
\pi(\theta|A) \propto \exp \left( -\frac{1}{2} \text{tr} (I_p - A)^{-1} \theta \theta^t \right) I(0_{p \times p} < A < I_p),
$$

(2.6)

where $I(\cdot)$ stands for the indicator function and $0_{p \times p} < A$ and $A < I_p$ mean that $A$ and $I_p - A$ are positive definite, respectively. Assume, in addition, that the second stage prior for
A, \pi(\Lambda), is orthogonally invariant, namely, \pi(\Lambda) = \pi(PAP^t) for P \in O_p. This assumption guarantees that the resulting Bayes estimator is equivariant with order-preserving property.

As the second stage prior \pi(\Lambda), we here focus on the Jeffreys-type prior of the form
\[
\pi(\Lambda) \propto |\Lambda|^{a/2-1} I(0_{p \times p} < \Lambda < I_p),
\]
where \(|\Lambda|\) means the determinant of \(\Lambda\) and \(a\) is a suitable constant. The hierarchical prior, given in (2.6) and (2.7), is regarded as natural multivariate extension of [14]'s prior for estimation of the normal mean vector. The density (2.7) is proper if \(a > 0\) and improper otherwise. In this subsection our interest is the admissibility and minimaxity of the resulting generalized Bayes estimator based on both (2.6) and (2.7).

Note that the priors (2.6) and (2.7) are motivated as follows: Suppose that \(\pi(\Lambda)\), \(\pi(\Theta)\), \(\pi(\Sigma)\), given \(\delta\), are motivated as follows: Suppose that
\[
\Lambda, \pi(\Lambda), \text{ and } \Theta, \pi(\Theta), \text{ and } \Sigma, \pi(\Sigma), \text{ are motivated as follows: Suppose that}
\]
\[
\pi(\Lambda) \propto |\Lambda|^{a/2-1} I(0_{p \times p} < \Lambda < I_p),
\]
where \(|\Lambda|\) means the determinant of \(\Lambda\) and \(a\) is a suitable constant. The hierarchical prior, given in (2.6) and (2.7), is regarded as natural multivariate extension of [14]'s prior for estimation of the normal mean vector. The density (2.7) is proper if \(a > 0\) and improper otherwise. In this subsection our interest is the admissibility and minimaxity of the resulting generalized Bayes estimator based on both (2.6) and (2.7).

Since the posterior densities of \(\Theta\) given \(\Lambda\) and \(X\) and of \(\Lambda\) given \(X\) with respect to the hierarchical prior given in (2.6) and (2.7) are, respectively,
\[
\pi(\Theta|X, \Lambda) \propto \exp \left( -\frac{1}{2} \text{tr} (I_p - \Lambda)^{-1} \{\Theta - (I_p - \Lambda)X\} \{\Theta - (I_p - \Lambda)X\}' \right),
\]
\[
\pi(\Lambda|X) \propto |\Lambda|^{(m+a)/2-1} \exp \left( -\frac{1}{2} \text{tr} \Lambda XX' \right) I(0_{p \times p} < \Lambda < I_p),
\]
the resulting Bayes estimator is represented by
\[
\delta^{BJ} = (I_p - U \Phi^{BJ}(L^2)U^t)X.
\]

It is important to note that
\[
\Phi^{BJ}(L^2) = \frac{\int_{0_{p \times p} < \Lambda < I_p} A |\Lambda|^{(m+a)/2-1} \exp(-1/2 \text{tr} \Lambda L^2) (d\Lambda)}{\int_{0_{p \times p} < \Lambda < I_p} |\Lambda|^{(m+a)/2-1} \exp(-1/2 \text{tr} \Lambda L^2) (d\Lambda)}
\]
is a diagonal matrix by equivariance of the Bayes estimators \(\delta^{BJ}\).

The following lemma, owing to Stein [13], is crucial for proving the minimaxity of the Bayes estimator \(\delta^{BJ}\).

**Lemma 2.2.** For \(i = 1, \ldots, p\), write \(\phi_i = \phi_i(F)\) and \(F = \text{diag}(f_1, \ldots, f_p) = L^2\). The risk of a shrinkage equivariant estimator \(\delta^s = UL(I_p - \Phi(F))V^t\) is evaluated by
\[
R(\delta^s, \Theta) = mp + E \left[ \sum_{i=1}^p \left( f_i \phi_i^2 - 2(m-p+1)\phi_i - 4 f_i \frac{\partial \phi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \phi_i - f_j \phi_j}{f_i - f_j} \right) \right]
\]
(2.10)
provided each expectation exists.
Proposition 2.1. If \(-m < a \leq m - 3p - 1\) and \(2m - 3p - 1 > 0\), then \(\delta^{BJ}\) is minimax. If \(0 < a \leq m - 3p - 1\) and \(m - 3p - 1 > 0\), then \(\delta^{BJ}\) is minimax and likewise admissible.

Proof. Recall that the prior distribution of \(\Lambda\) is proper for \(a > 0\) and note that if the proper Bayes estimator has smaller risk than the minimax risk \(mp\) then it is admissible. Thus, using these facts and the first part of Proposition, namely, the condition for minimaxity of \(\delta^{BJ}\), yields the proof for the second part of Proposition.

To prove the first part of Proposition, let \(\Phi^{BJ}(F) = \text{diag}(\phi^{BJ}_1, \ldots, \phi^{BJ}_p)\), where \(F = L^2\) and

\[
\phi^{BJ}_i = \frac{\int_{0 < \lambda < 1} \lambda_i |A|^{(m+a)/2-1} \exp\left(-\frac{1}{2}\text{tr} A F (dA)\right)}{\int_{0 < \lambda < 1} |A|^{(m+a)/2-1} \exp\left(-\frac{1}{2}\text{tr} A F (dA)\right)}
\]

for \(A = (\lambda_{ij})\).

Now, let us consider the change of variables \(\lambda_{ij} = \gamma_{ij} \sqrt{\lambda_{ii} \lambda_{jj}}\) for \(i < j\). The Jacobian of this transformation is given by

\[
J[(\lambda_{11}, \ldots, \lambda_{pp}, \lambda_{12}, \ldots, \lambda_{p-1,p}) \rightarrow (\lambda_{11}, \ldots, \lambda_{pp}, \gamma_{12}, \ldots, \gamma_{p-1,p})] = \prod_{i=1}^{p} \lambda_{ii}^{(p-1)/2}.
\]

It also follows that \(|A| = |\Gamma| |\prod_{i=1}^{p} \lambda_{ii}|\), where \(\Gamma = (\gamma_{ij})\) is \(p \times p\) positive definite matrix with \(\gamma_{ii} = 1\) for \(i = 1, \ldots, p\). Denote \((d\Gamma) = \sum_{i<j} d\gamma_{ij}\) and \((d\lambda) = \sum_{i=1}^{p} d\lambda_{ii}\) and use the notation \(0 < \lambda < 1\) which means that \(0 < \lambda_{ii} < 1\) for \(i = 1, \ldots, p\). Then \(\phi^{BJ}_i\) can be written as

\[
\phi^{BJ}_i = \frac{\int_{0 < \lambda < 1} \lambda_i |\Gamma|^{(m+a)/2-1} (d\Gamma)}{\int_{0 < \lambda < 1} |\Gamma|^{(m+a)/2-1} (d\Gamma) \prod_{i=1}^{p} \lambda_{ii}^{(m+a+p-3)/2} \exp(-\lambda_{ii} f_i/2)} \left(\sum_{i=1}^{p} \lambda_{ii}^{(m+a+p-3)/2} \exp(-\lambda_{ii} f_i/2) \right) (d\lambda)
\]

for

\[
m + a > 0,
\]

which is the condition for finiteness of \(\int_{0 < \lambda < 1} \lambda_i |\Gamma|^{(m+a)/2-1} (d\Gamma)\). Moreover, the integration by parts yields

\[
\phi^{BJ}_i = \frac{1}{f_i} \left( m + a + p - 1 - \frac{2 \exp(-f_i/2)}{\int_{0 < \lambda_{ii} < 1} \lambda_{ii}^{(m+a+p-3)/2} \exp(-\lambda_{ii} f_i/2) d\lambda_{ii}} \right) = \frac{\xi_i}{f_i}.
\]

Thus, using Lemma 2.2 gives

\[
R(\delta^{BJ}, \Theta) = mp + E \left[ \sum_{i=1}^{p} \frac{1}{f_i} \left( \xi_i^2 - 2(\xi_i - f_i) + 4 \frac{\partial \xi_i}{\partial f_i} \right) - 4 \sum_{i=1}^{p} \sum_{j>i} \frac{\xi_i - \xi_j}{f_i - f_j} \right].
\]
It is here obvious that
\[ \xi_i = m + a + p - 1 - \frac{2}{\int_0^1 \lambda_i^{(m+a+p-3)/2} \exp\{1/2(1-\lambda_i)f_i\}d\lambda_i} \]
is nondecreasing in \( f_i \) and that \( 0 < \xi_i \leq m + a + p - 1 \). Also, note that \( \xi_1 > \cdots > \xi_p \) and \( f_1 > \cdots > f_p \). Therefore, the risk of \( \delta^{BJ} \) is smaller than the minimax risk \( mp \) if
\[ 0 < m + a + p - 1 \leq 2(m - p - 1). \] (2.12)
Combining (2.11) and (2.12) provides the condition for minimaxity of \( \delta^{BJ} \). \( \blacksquare \)

Proposition 2.1 for \( p = 1 \) is essentially equivalent to [14]'s result. For \( a \leq -p + 1 - 2/p \), the inadmissibility of the generalized Bayes estimator \( \delta^{BJ} \) will be discussed in the next section (see Corollary 3.4). Perhaps \( \delta^{BJ} \) for \( -p + 1 - 2/p < a \leq 0 \) is admissible, but it seems hard to prove.

3. Improvement upon nonorder-preserving and overshrinkage estimators

3.1. General results

The order-preserving and nonnegativity properties of the Stein estimator \( \delta^{ST} \) do not necessarily hold since \( l_1 > \cdots > l_p \) and \( d_1 > \cdots > d_p \). For the equivariant estimator \( \delta = U \Psi(L)V \) with \( \Psi(L) = \text{diag} (\psi_1(L), \ldots, \psi_p(L)) \), it is likely reasonable to impose the ordering \( \psi_1(L) \geq \cdots \geq \psi_p(L) \) and the nonnegativity of the \( \phi_i(L) \)'s since the singular values keep the order and positivity \( l_1 > \cdots > l_p > 0 \). In this subsection it is shown that a nonorder-preserving equivariant estimator is dominated by the corresponding estimator modifying both nonorder-preserving and negativity properties.

First, let us derive the joint density of \( L, U \) and \( V \). The density of \( X = (x_{ij}) \) is given by
\[ (2\pi)^{-mp/2} \exp(-\|X - \Theta\|^2/2) (dX), \]
where \( (dX) = \bigwedge_{i=1}^p \bigwedge_{j=1}^m dx_{ij} \). Let \( U = (u_1, \ldots, u_p) \in \mathcal{O}_p, V = (v_1, \ldots, v_p) \in \mathcal{V}_{m,p} \). Uhlig [16] showed that the Jacobian of transformation by the singular value decomposition \( X = ULV \) is written by
\[ (dX) = 2^{-p} \prod_{i=1}^p l_i^{-m-p} \prod_{i<j} (l_i^2 - l_j^2) (dU) (dL) (dV), \]
where
\[ (dL) = \bigwedge_{i=1}^p dl_i, \quad (dU) = \bigwedge_{i<j} u_i^j du_i \quad \text{and} \quad (dV) = \bigwedge_{i=1}^p \bigwedge_{j=1}^m v_i^j dv_i \]
with \( (V : V_0) = (v_1, \ldots, v_p : v_{p+1}, \ldots, v_m) \in \mathcal{O}_m \) being a function of \( V \). See also [6]. Hence, the joint density of \( L, U \) and \( V \) can immediately be expressed as
\[ 2^{-p} (2\pi)^{-mp/2} \exp(-\|ULV - \Theta\|^2/2) \prod_{i=1}^p l_i^{-m-p} \prod_{i<j} (l_i^2 - l_j^2) (dL) (dU) (dV) = k_{m,p}(\Theta) \exp \left( -\frac{1}{2} \sum_{i=1}^p l_i^2 + \sum_{i=1}^p l_i a_i \right) \prod_{i=1}^p l_i^{-m-p} \prod_{i<j} (l_i^2 - l_j^2) (dL) (dU) (dV), \]
where \( k_{m,p}(\Theta) = \frac{2^{-p}(2\pi)^{-mp/2}}{\|\Theta\|/2} \) and
\[
a_i = (U^t \Theta V)_{ii}
\]
for \( i = 1, \ldots, p \).

The following lemma holds the key to proving the main result of this subsection:

**Lemma 3.1.** For \( i = 1, \ldots, p \), denote by \( E[a_i|L] \) the conditional expectation of \( a_i \) given \( L \). Then \( E[a_i|L] \geq \cdots \geq E[a_p|L] \geq 0 \).

**Proof.** This proof will be done by the similar way in [12] or Lemma 2.1 of the preceding section. It is sufficient to show that for \( 1 \leq i < j \leq p \)
\[
\int \int \mathcal{O}_p \times \mathcal{V}_{m,p} (a_i - a_j) \exp \left( \sum_{k=1}^{p} l_k a_k \right) (dU)(dV) \geq 0
\]
and that
\[
\int \int \mathcal{O}_p \times \mathcal{V}_{m,p} a_p \exp \left( \sum_{k=1}^{p} l_k a_k \right) (dU)(dV) \geq 0.
\]

The distributions of \( U \) and \( V \) are, respectively, invariant under the orthogonal transformations \( U \rightarrow UP \) and \( V \rightarrow VP \) for \( P \in \mathcal{O}_p \) ([10], p.p.69). Hence, we make the orthogonal transformations \( (U, V) \rightarrow (UO_{ij}, VO_{ij}) \) and \( V \rightarrow VP \) where \( O_{ij} \) and \( P_p \) are given in the proof of Lemma 2.1, and use the same arguments as in the proof of Lemma 2.1 to obtain the desired results (3.1) and (3.2). \( \blacksquare \)

**Proposition 3.1.** Let \( \Psi = \text{diag}(\psi_1, \ldots, \psi_p) \) and \( \Psi_0 = \text{diag}(\psi_{01}, \ldots, \psi_{0p}) \), where \( \psi_i \)'s and \( \psi_{0i} \)'s are functions of \( L \), respectively. Define two estimators of \( \Theta \) as \( \delta = U\Psi \Psi^t \) and \( \delta_0 = U\Psi_0 \Psi_0^t \). Assume that \( \sum_{i=1}^{p} \psi_{0i}^2 \leq \sum_{i=1}^{p} \psi_i^2 \) and that \( \psi_{0i} \)'s weakly majorize \( \psi_i \)'s, that is, for \( j = 1, \ldots, p \)
\[
\sum_{i=1}^{j} \psi_{0i} \geq \sum_{i=1}^{j} \psi_i.
\]
If \( \Pr(\Psi \neq \Psi_0) > 0 \), then \( \delta_0 \) dominates \( \delta \) under the loss (1.2).

**Proof.** The risk difference of \( \delta_0 \) and \( \delta \) is expressed by
\[
R(\delta_0, \Theta) - R(\delta, \Theta) = E[\|U\Psi_0 \Psi^t - \Theta\|^2 - \|U\Psi \Psi^t - \Theta\|^2] = E \left[ \sum_{i=1}^{p} (\psi_{0i}^2 - \psi_i^2 - 2(\psi_{0i} - \psi_i)E[a_i|L]) \right],
\]
which is nonpositive if \( \sum_{i=1}^{p} (\psi_{0i}^2 - \psi_i^2) \leq 0 \) and \( \sum_{i=1}^{p} (\psi_{0i} - \psi_i)E[a_i|L] \geq 0 \). By the same argument as in [12], the Abel summation gives that
\[
\sum_{i=1}^{p} (\psi_{0i} - \psi_i)E[a_i|L] = (\psi_{01} - \psi_1)E[a_1 - a_2|L]
\]
\[
+ (\psi_{01} + \psi_{02} - \psi_1 - \psi_2)E[a_2 - a_3|L] + \cdots
\]
\[
+ (\psi_{01} + \cdots + \psi_{0p-1} - \psi_1 - \cdots - \psi_{p-1})E[a_{p-1} - a_p|L]
\]
\[
+ (\psi_{01} + \cdots + \psi_{0p} - \psi_1 - \cdots - \psi_p)E[a_p|L],
\]
which is nonnegative from Lemma 3.1 and the assumption. Thus the proof is complete. \( \blacksquare \)
Proposition 3.1 allows us to provide some methods for modifying nonorder-preserving and overshrinkage estimators $\delta = U \Psi^s V^t$ with $\Psi^s = \Psi^s(L) = \text{diag}(\psi^s_1, \ldots, \psi^s_p)$. Let $\Psi^s_+$ be diagonal matrix whose $i$-th diagonal element is $\max(0, \psi^s_i)$. Also, let $\Psi^s_{O+}$ be diagonal matrix with the $i$-th diagonal element being $\max(0, \psi^s_{(i)})$ where $\psi^s_{(i)}$’s are order statistics of $\psi^s_i$’s, that is, $\psi^s_{(i)}$ is the $i$-th largest value of $\psi^s_i$’s. It is clear that $\Psi^s_+$ and $\Psi^s_{O+}$ satisfy the assumptions over $\Psi^s$ in Proposition 3.1.

**Corollary 3.1.** Assume that $\delta^s = U \Psi^s V^t$ is nonorder-preserving and over-shrinking, namely, $\Pr(\psi^s_1 \geq \cdots \geq \psi^s_p \geq 0) < 1$. Then $\delta^s$ is dominated by $\delta^s_+ = U \Psi^s_+ V^t$ or $\delta^s_{O+} = U \Psi^s_{O+} V^t$ under the loss (1.2).

As modification technique of the order, the isotonic regression is also utilized. Let $\Psi^s_{I+} = \text{diag}(\max(0, \psi^s_1), \ldots, \max(0, \psi^s_p))$, where $\psi^s_i$’s be the unique solution of

$$
\min_{\beta_1 \geq \cdots \geq \beta_p} \sum_{i=1}^p (\beta_i - \psi^s_i)^2 = \sum_{i=1}^p (\tilde{\psi}_i - \psi^s_i)^2.
$$

It is noted that $\tilde{\psi}_i$’s are obtained by the isotonic regression and for the computational algorithm of the solution, see [11]. Using the similar arguments of Sheena and Takeamura [12], we can easily show that the estimator with $\Psi^s_{I+}$ dominates its original nonorder-preserving estimator.

**Corollary 3.2.** Suppose $\delta^s = U \Psi^s V^t$ is nonorder-preserving and overshrinking. Then $\delta^s_{I+} = U \Psi^s_{I+} V^t$ dominates $\delta^s$ under the loss (1.2).

3.2. New minimax estimators

Applying Proposition 3.1, we are able to obtain an order-preserving and positive-part estimator improving upon its original nonorder-preserving estimator and it is, however, difficult to look for new minimax estimators different from $\delta^{EM}$, $\delta^{ST}$ and $\delta^{BJ}$. To do this, Lemma 2.2 is very instrumental tool. The proofs of the following results are deferred to the next subsection.

First, let us consider further improvement upon shrinkage estimators $\delta^s = U L (I_p - \Phi(F)) V^t$ where $L = \text{diag}(f_1, \ldots, f_p) = L^2$ and $\Phi(F) = \text{diag}(\phi_1(F), \ldots, \phi_p(F))$. Their improved estimators, considered here, have the form

$$
\delta^m = \delta^s - (c/\text{tr} XX^t)X = UL(I_p - \Phi^m(F)) V^t,
$$

where $\Phi^m(F) = \Phi(F) + (c/\text{tr} F) I_p$ for a suitable constant $c$. This type of improved estimators is found in [13, 17, 18]. Then the following proposition holds:

**Proposition 3.2.** Assume that $A(\Phi(F)) = \text{tr} F \Phi(F) = \sum_{i=1}^p f_i \phi_i(F) \leq D$, where $D$ is a constant and less than $mp - 2$. If $0 < c \leq 2(mp - 2 - D)$, then $\delta^m$ dominates $\delta^s$ relative to the loss (1.2).

For instance, the improvement upon the Stein estimator $\delta^{ST}$ is done as follows: Since $\phi^{ST}_i = (m + p - 2i - 1)/f_i$ for $i = 1, \ldots, p$, we obtain $A(\Phi^{ST}(F)) = \sum_{i=1}^p (m + p - 2i - 1) = p(m - 2)$. Thus $\delta^{ST}$ is dominated by $\delta^{MST} = \delta^{ST} - (c/\text{tr} XX^t)X$ for $0 < c \leq 4(p - 1)$.

We next give two types of new minimax estimators of $\Theta$.

(i) For $i = 1, \ldots, p$, let $\phi^{N1}_i(F) = b_i f_i / T_1$, where $b_i$ is a constant and $T_1 = \sum_{i=1}^p f_i^2$. Define $\delta^{N1} = UL(I_p - \Phi^{N1}(F)) V^t$. 
(ii) For $i = 1, \ldots, p$, let $\phi_i^{N2}(L) = c_i f_i^{-2}/T_2$, where $c_i$ is a constant and $T_2 = \sum_{i=1}^{p} f_i^{-1}$.
Define $\delta^{N2} = UL(I_p - \Phi^{N2}(F))V^t$.

**Lemma 3.2.** (i) Suppose $2(m+p-3) \geq b_1 \geq \cdots \geq b_p > 0$ for $m+p-3 > 0$. Then $\delta^{N1}$ is minimax relative to the loss (1.2).
(ii) Suppose $c_1 \geq \cdots \geq c_p$ and that $0 < c_i \leq 2(m-p+1-2i)$ for $i = 1, \ldots, p$ with $m-3p-1 > 0$. Then $\delta^{N2}$ is minimax relative to the loss (1.2).

The minimaxity of $\delta^{N1}$ holds even if $m = p \geq 2$. Meanwhile the condition that $\delta^{EM}$ and $\delta^{ST}$ are minimax is that $m-p-1 > 0$ and, also, $\delta^{N2}$ is minimax for $m-3p-1 > 0$. Proposition 3.2 also gives us its corollary for refinement on $\delta^{N1}$ and $\delta^{N2}$.

**Corollary 3.3.** In the problem of estimating the mean matrix $\Theta$ under the loss (1.2), we have the following:

(i) If $0 < c \leq 2(mp - b_1)$ and $b_1 < mp - 2$, then $\delta^{MN1} = \delta^{N1} - (c/\text{tr} XX^t)X$ dominates $\delta^{N1}$.
(ii) If $0 < c \leq 2(mp - c_1)$ and $c_1 < mp - 2$, then $\delta^{MN2} = \delta^{N2} - (c/\text{tr} XX^t)X$ dominates $\delta^{N2}$.

Next, we reconsider the generalized Bayes estimator $\delta^{BJ}$ given in (2.9). Although $\delta^{BJ}$ for $a \leq -p + 1 - 2/p$ is order-preserving and nonovershrinking, Proposition 3.2 makes it possible to prove its inadmissibility.

**Corollary 3.4.** If $-m < a \leq -p + 1 - 2/p$ and $0 < c < 2p(1-p-a) - 4$, then $\delta^{BJ}$ is dominated by $\delta^{MBJ} = \delta^{BJ} - (c/\text{tr} XX^t)X$ relative to the loss (1.2).

It is noted that $\delta^{MBJ}$ does not satisfy the nonnegativity property any longer and itself is improved upon by its positive-part estimator.

Since the constant $c$ of $\Psi^m$ is positive, it is conceivable that $\delta^{EM}$, $\delta^{ST}$, $\delta^{N1}$ and $\delta^{N2}$ are not enough shrinking toward the zero matrix. In general, there is no guarantee that each element of $\Psi^m$ is order-preserving and nonnegative. Therefore, modifying these undesirable properties as in the preceding subsection, we can obtain an estimator with the natural order and nonnegativity properties. Finally, we conclude this subsection by giving a general minimaxity result with respect to the shrinkage estimators.

**Proposition 3.3.** If a nonorder-preserving and overshrinkage estimator $\delta^s = U \Psi^s V^t$ is minimax relative to the loss (1.2), then $\delta^s_+ = U \Psi^s_+ V^t$, $\delta^s_o+ = U \Psi^s_o+ V^t$ and $\delta^s_{1+} = U \Psi^s_{1+} V^t$ are also minimax.

### 3.3. Proofs

Write the expression in the large bracket of (2.10) as $\hat{\Delta}(\Phi)$, that is,

$$
\hat{\Delta}(\Phi) = \sum_{i=1}^{p} \left\{ f_i \phi_i^2 - 2(m-p+1)\phi_i - 4 f_i \frac{\partial \phi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \phi_i - f_j \phi_j}{f_i - f_j} \right\},
$$

(3.3)

where $f_i = \lambda_i^2$. If $\hat{\Delta}(\Phi)$ is nonpositive, then the shrinkage estimator with $\Phi = \Phi(F)$ is minimax relative to the loss (1.2).
Proof of Proposition 3.2. Note that the risk difference $R(\delta^m, \Theta) - R(\delta^s, \Theta)$ is non-positive if $\hat{\Delta}(\Phi^m) - \hat{\Delta}(\Phi)$ is nonpositive. Now, the difference of $\hat{\Delta}(\Phi^m)$ and $\hat{\Delta}(\Phi)$ is evaluated as

$$
\hat{\Delta}(\Phi^m) - \hat{\Delta}(\Phi) = \sum_{i=1}^{p} \left\{ f_i \left( \frac{c^2}{t^2} + \frac{2c\phi_i}{t} \right) - 2(m - p + 1) \frac{c}{t} - 4cf_i \frac{\partial}{\partial f_i} \frac{1}{t} - \frac{4}{t} \sum_{j>i} \right\}
$$

where $t = \text{tr} F$. Hence if $\sum_{i=1}^{p} f_i \phi_i \leq D$ then we have the desired result. ■

Proof of Lemma 3.2. For the proof of part (i), using (3.3) gives

$$
\hat{\Delta}(\Phi^{N1}) = \sum_{i=1}^{p} \left\{ \frac{b_i f_i^2}{T_1^2} - 4f_i \frac{\partial}{\partial f_i} \frac{b_i f_i}{T_1} - 2(m - p + 1) \frac{b_i f_i}{T_1} - \frac{4}{T_1} \sum_{j>i} \frac{b_i f_i^2 - b_j f_j^2}{f_i - f_j} \right\}
$$

It is here observed that

$$
\sum_{i=1}^{p} \sum_{j>i} \frac{b_i f_i^2 - b_j f_j^2}{f_i - f_j} = \sum_{i=1}^{p} \sum_{j>i} \frac{b_i(f_i^2 - f_j^2) + (b_i - b_j)f_j^2}{f_i - f_j}
$$

$$
= \sum_{i=1}^{p} \sum_{j>i} b_i(f_i + f_j) + \sum_{i=1}^{p} \sum_{j>i} \frac{b_i - b_j}{f_i - f_j} f_j^2
$$

and, additionally, that

$$
\sum_{i=1}^{p} \sum_{j>i} b_i(f_i + f_j) = \sum_{i=1}^{p} \left\{ b_i f_i \sum_{j>i} 1 + \sum_{j>i} b_i f_j \right\}
$$

$$
\geq \sum_{i=1}^{p} \left\{ (p - i)b_i f_i + \sum_{j>i} b_j f_j \right\} = \sum_{i=1}^{p} (p - 1)b_i f_i.
$$

The above inequality leads us to

$$
\hat{\Delta}(\Phi^{N1}) \leq \sum_{i=1}^{p} \left\{ \frac{f_i^3}{T_1^2} (b_i^2 + 8b_i) - 2(m + p + 1) \frac{b_i f_i}{T_1} - \frac{4}{T_1} \sum_{j>i} \frac{b_i - b_j}{f_i - f_j} f_j^2 \right\}
$$

$$
\leq \sum_{i=1}^{p} \frac{f_i}{T_1} \left\{ b_i^2 - 2(m + p - 3)b_i \right\} - \frac{4}{T_1} \sum_{i=1}^{p} \sum_{j>i} \frac{b_i - b_j}{f_i - f_j} f_j^2,
$$

where the last inequality follows from the fact that $f_i^2 / T_1 \leq 1$. Thus the proof of part (i) is complete. ■
By replacing $\Phi$ with $\Phi_{N_2}^2$ in (3.3) and some straightforward calculations, we obtain
\[
\Delta(\Phi_{N_2}^2) = \sum_{i=1}^{p} \frac{f_i^{-2}}{T_2} \left\{ \frac{f_i^{-1}}{T_2} c_i^2 - 2(m - p - 3)c_i - \frac{4c_i f_i^{-1}}{T_2} \right\}.
\]
\[
- \frac{4}{T_2} \sum_{i=1}^{p} \sum_{j>i} c_i f_i^{-1} - c_j f_j^{-1} \frac{f_i - f_j}{f_i f_j}.
\]
It is then seen that
\[
\sum_{i=1}^{p} \sum_{j>i} c_i f_i^{-1} - c_j f_j^{-1} \frac{f_i - f_j}{f_i f_j} = \sum_{i=1}^{p} \sum_{j>i} (c_i - c_j) f_j f_i (f_i - f_j)
\]
\[
= \sum_{i=1}^{p} \sum_{j>i} (c_i - c_j) f_i f_j - \sum_{i=1}^{p} \sum_{j>i} \frac{c_j}{f_i f_j},
\]
and, furthermore, that
\[
\sum_{i=1}^{p} \sum_{j>i} \frac{c_j}{f_i f_j} \leq \sum_{i=1}^{p} \sum_{j>i} \frac{c_j}{f_j^2} = \sum_{i=1}^{p} (i - 1) \frac{c_i}{f_i^2}.
\]
Combining (3.4) and (3.5) and the assumption gives
\[
\sum_{i=1}^{p} \sum_{j>i} \frac{c_i f_i^{-1} - c_j f_j^{-1}}{f_i - f_j} \geq - \sum_{i=1}^{p} (i - 1) \frac{c_i}{f_i^2},
\]
which yields that
\[
\Delta(\Phi_{N_2}^2) \leq \sum_{i=1}^{p} \frac{f_i^{-2}}{T_2} \left\{ \frac{f_i^{-1}}{T_2} c_i^2 - 2(m - p - 1 - 2i)c_i - \frac{4c_i f_i^{-1}}{T_2} \right\}.
\]
If $m - 3p - 1 > 0$, then the fact that $f_i^{-1}/T_2 \leq 1$ yields that
\[
\Delta(\Phi_{N_2}^2) \leq \sum_{i=1}^{p} \frac{f_i^{-3}}{T_2} \{c_i^2 - 2(m - p + 1 - 2i)c_i\},
\]
which completes the proof of part (ii). \[\square\]

Proof of Corollary 3.3. This is trivial since $b_1 \geq \cdots \geq b_p$ and $c_1 \geq \cdots \geq c_p$. \[\square\]

Proof of Corollary 3.4. This proof can be verified by the fact that $A(F) = \sum_{i=1}^{p} \xi_i \leq p(m + a + p - 1)$. \[\square\]

Acknowledgments

The author thanks Professor Tatsuya Kubokawa for his helpful comments to improve this paper. The author is also grateful to two reviewers for their careful reading and valuable suggestions.
References