CIRJE-F-926

A Unified Approach to Estimating a Normal Mean Matrix in High and Low Dimensions

Hisayuki Tsukuma
Toho University

Tatsuya Kubokawa
The University of Tokyo

March 2014; Revised in October 2014

A Unified Approach to Estimating a Normal Mean Matrix in High and Low Dimensions

Hisayuki Tsukuma* and Tatsuya Kubokawa†

October 13, 2014

Abstract

This paper addresses the problem of estimating the normal mean matrix with an unknown covariance matrix. Motivated by an empirical Bayes method, we suggest a unified form of the Efron-Morris type estimators based on the Moore-Penrose inverse. This form not only can be defined for any dimension and any sample size, but also can contain the Efron-Morris type or Baranchik type estimators suggested so far in the literature. Also, the unified form suggests a general class of shrinkage estimators. For shrinkage estimators within the general class, a unified expression of unbiased estimators of the risk functions is derived regardless of the dimension of covariance matrix and the size of the mean matrix. An analytical dominance result is provided for a positive-part rule of the shrinkage estimators.

AMS 2010 subject classifications: Primary 62C10; secondary 62F10.

Key words and phrases: Eforn-Morris estimator, empirical Bayes procedure, high dimension, invariant loss, matrix mean, Moore-Penrose inverse, shrinkage estimator, statistical decision theory.

1 Introduction

Statistical inference with high dimension has received much attention in recent years, because statistical analysis of high-dimensional data has been requested in many research areas such as genomics, remote sensing, telecommunication, atmospheric science, financial engineering, and others. Such high-dimensional data are generally hard to handle, and ordinary or traditional methods are frequently inapplicable. This has inspired statisticians to develop new research areas in high dimension from both theoretical and practical

*Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan, E-Mail: tsukuma@med.toho-u.ac.jp

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, E-Mail: tatsuya@e.u-tokyo.ac.jp

aspects. Most interests have been in development of efficient algorithm for statistical inference and in derivation of asymptotic properties with the dimension going to infinity. From a decision-theoretic point of view, however, there does not exist much literature in high-dimensional problems except for Chételat and Wells (2012), who established the inadmissibility of the maximum likelihood estimator (MLE) for a large dimensional and small sample normal model. In this paper, we extend their result to the framework of estimating a mean matrix and we establish a unified theory for improvement on the MLE in both cases of high and low dimensions.

To explain the subjects addressed here, we begin by describing the canonical model and the estimation problem. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)^t$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^t$ be, respectively, $m \times p$ and $n \times p$ random matrices, where $\boldsymbol{X}_i$'s and $\boldsymbol{Y}_i$'s are mutually and independently distributed as

$$\begin{aligned} \boldsymbol{X}_i &\sim \mathcal{N}_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}), \quad i = 1, \ldots, m, \\ \boldsymbol{Y}_j &\sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Sigma}), \quad j = 1, \ldots, n. \end{aligned} \tag{1.1}$$

Suppose that $\boldsymbol{\theta}_i$'s are unknown mean vectors and that $\boldsymbol{\Sigma}$ is an unknown positive definite matrix. It is noted that the model (1.1) is a canonical form of a multivariate linear regression model although the details are omitted here. The problem we consider in this paper is the estimation of the mean matrix $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)^t$ relative to the invariant quadratic loss

$$L(\boldsymbol{\delta}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) = \operatorname{tr}(\boldsymbol{\delta} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\delta} - \boldsymbol{\Theta})^t, \tag{1.2}$$

where $\boldsymbol{\delta}$ is an estimator made from $\boldsymbol{X}$ and $\boldsymbol{S}$.

The MLE of $\boldsymbol{\Theta}$ is $\boldsymbol{\delta}^{ML} = \boldsymbol{X}$, which is a minimax estimator with the constant risk mp. When $n \geq p$, it is known that $\boldsymbol{\delta}^{ML}$ is improved on by the Efron-Morris (1972) type estimator

$$\boldsymbol{\delta}^{EMK} = \begin{cases} \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1}\boldsymbol{X} & \text{if } p \geq m, \\ \boldsymbol{X} - c\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S} & \text{if } m > p, \end{cases}$$

where $\boldsymbol{S} = \boldsymbol{Y}^t\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{Y}_i\boldsymbol{Y}_i^t$, and c is a suitable constant. Konno (1991, 1992) derived conditions on c for the improvement. When $p > n$, however, this estimator is not available because the inverse $\boldsymbol{S}^{-1}$ does not exist. A possible alternative is the Moore-Penrose inverse $\boldsymbol{S}^+$ which will be defined in the beginning of Section 2. In the case of $m = 1$, Chételat and Wells (2012) suggested the shrinkage estimator

$$\boldsymbol{\delta}^{CW} = \boldsymbol{X} - \frac{c}{\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t}\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$$

and provided a condition on c for $\boldsymbol{\delta}^{CW}$ to dominate $\boldsymbol{\delta}^{ML}$.

An interesting issue is how to extend the Chételat-Wells estimator $\boldsymbol{\delta}^{CW}$ to the framework of estimation of the mean matrix $\boldsymbol{\Theta}$. Especially, the Efron-Morris type estimators seem to take various variants which depend on orderings among m, p and n. One of interesting results provided in this paper is that we can develop a unified form for the Efron-Morris type estimators, given by

$$\boldsymbol{\delta}^{EM} = \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{S}^{+}\boldsymbol{X}^{t})^{+}\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^{+}.$$

As explained in Section 2, this estimator can be defined for all the positive integers of m, p and n as well as this expression includes $\boldsymbol{\delta}^{EMK}$ and $\boldsymbol{\delta}^{CW}$ as special cases. Also this expression suggests us to consider a general class of shrinkage estimators in Section 3. In this paper, we derive a unified expression of an unbiased estimator of the risk function for shrinkage estimators within the general class.

The paper is organized as follows. In Section 2, we introduce shrinkage estimators of $\boldsymbol{\Theta}$ based on a motivation from an empirical Bayes method, and we provide the unified form of the Efron-Morris type estimator. This expression not only contains the Efron-Morris or Baranchik type estimators suggested so far in the literature, but also provide various forms corresponding to ordering of m, p and n. In Section 3, we consider a general class of shrinkage estimators. A unified expression is developed in Section 4 for the risk functions of the general shrinkage estimators. It is noted that the unified expression gives an unbiased estimator of the risk difference. As specific examples of shrinkage estimators, we treat the modified Efron-Morris type estimators and the modified Stein type estimators, and we get conditions for their improvement from the unified expression. Section 5 provides analytical and numerical dominance results that positive-part estimators improve the corresponding shrinkage estimators. Some technical proofs are given in Section 6.

2 A Bayesian Motivation

We begin by describing basic and useful properties of the Moore-Penrose inverse. For any matrix $\boldsymbol{A}$, the Moore-Penrose inverse of $\boldsymbol{A}$ is written by $\boldsymbol{A}^{+}$ if $\boldsymbol{A}^{+}$ satisfies (i) $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{A}$, (ii) $\boldsymbol{A}^{+}\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{A}^{+}$, (iii) $(\boldsymbol{A}\boldsymbol{A}^{+})^{t} = \boldsymbol{A}\boldsymbol{A}^{+}$, and (iv) $(\boldsymbol{A}^{+}\boldsymbol{A})^{t} = \boldsymbol{A}^{+}\boldsymbol{A}$. The Moore-Penrose inverse $\boldsymbol{A}^{+}$ has the following properties: (1) $\boldsymbol{A}^{+}$ uniquely exists; (2) $(\boldsymbol{A}^{+})^{t} = (\boldsymbol{A}^{t})^{+}$; (3) $\boldsymbol{A}^{+} = \boldsymbol{A}^{-1}$ for a nonsingular matrix $\boldsymbol{A}$.

Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be $r \times p$ matrices of full row rank. We then have (1) $\boldsymbol{B}^{+} = \boldsymbol{B}^{t}(\boldsymbol{B}\boldsymbol{B}^{t})^{-1}$, (2) $\boldsymbol{B}\boldsymbol{B}^{+} = \boldsymbol{I}_{r}$, (3) $\boldsymbol{B}^{+}\boldsymbol{B}$ is idempotent, (4) $(\boldsymbol{B}^{t}\boldsymbol{C})^{+} = \boldsymbol{C}^{+}(\boldsymbol{B}^{t})^{+} = \boldsymbol{C}^{t}(\boldsymbol{C}\boldsymbol{C}^{t})^{-1}(\boldsymbol{B}\boldsymbol{B}^{t})^{-1}\boldsymbol{B}$. Further, for an $r \times r$ nonsingular matrix $\boldsymbol{A}$ and an $r \times q$ matrix $\boldsymbol{B}$ of full row rank, we can easily show that $(\boldsymbol{B}^{t}\boldsymbol{A}\boldsymbol{B})^{+} = \boldsymbol{B}^{+}\boldsymbol{A}^{-1}(\boldsymbol{B}^{t})^{+}$.

Based on the properties of the Moore-Penrose inverse, we give a unified form of empirical Bayes estimators. Using a similar argument as in Tsukuma and Kubokawa (2007), we can show that in the case of known $\boldsymbol{\Sigma}$, the empirical Bayes estimator of $\boldsymbol{\Theta}$ is given by

$$\boldsymbol{\delta}^{B} = \begin{cases} \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^{t})^{-1}\boldsymbol{X} & \text{for } p \geq m, \\ \boldsymbol{X} - c\boldsymbol{X}(\boldsymbol{X}^{t}\boldsymbol{X})^{-1}\boldsymbol{\Sigma} & \text{for } m > p, \end{cases}$$

for a suitable constant c. Here it is observed that, for $m > p$,

$$(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^+ = (\boldsymbol{X}^t)^+\boldsymbol{\Sigma}\boldsymbol{X}^+ = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{\Sigma}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t,$$

which yields that $(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^+\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{\Sigma}$. Hence, both cases $p \geq m$ and $m > p$ for the empirical Bayes estimator $\boldsymbol{\delta}^B$ can be unified by

$$\boldsymbol{\delta}^B = \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^+\boldsymbol{X}.$$

Since $\boldsymbol{\Sigma}^{-1}$ is unknown, we need to estimate it. In the case of $n \geq p$, $\boldsymbol{\Sigma}^{-1}$ is estimated by $n\boldsymbol{S}^{-1}$, so that we get the Efron-Morris type empirical Bayes estimator

$$\boldsymbol{\delta}^{EM} = \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^+\boldsymbol{X}. \tag{2.1}$$

The dominance properties of this estimator have been studied by Konno (1990, 1991, 1992).

In the case of $p > n$, the rank of $\boldsymbol{S}$ is deficient and its inverse does not exist. Therefore, we here estimate $\boldsymbol{\Sigma}^{-1}$ via $n\boldsymbol{S}^+$, where $\boldsymbol{S}^+$ is the Moore-Penrose inverse of $\boldsymbol{S}$.

The matrix $\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t$ is nonsingular for $p > n \geq m$, while it is singular for $p \geq m > n$. Taking the shrinkage estimator suggested by Chételat and Wells (2012) into account, we can suggest the Efron-Morris type shrinkage estimator given by

$$\boldsymbol{\delta}^{EM} = \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+, \tag{2.2}$$

for any set of (m, p, n). This gives a unified form of the Efron-Morris type estimator for any positive integers p, m and n. In fact, $(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+$ can be rewritten as

$$(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+ = \begin{cases} (\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1} & \text{for } n \geq p \geq m, \\ (\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^{-1} & \text{for } p > n \geq m, \\ (\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^+ & \text{for } n \geq m > p,\ m > n \geq p, \end{cases}$$

and the corresponding Efron-Morris estimators are provided. Especially, in the case that $m > p$, $\boldsymbol{\delta}^{EM}$ is expressed in the following proposition which is an extension of Konno's (1992) class for $m > p$.

Proposition 2.1 *In the case of $m > p$, the Efron-Morris type estimator given in* (2.2) *is expressed as*

$$\boldsymbol{\delta}^{EM} = \begin{cases} \boldsymbol{X} - c\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S} & \text{for } n \geq m > p,\ m > n \geq p, \\ \boldsymbol{X} - c\boldsymbol{X}(\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+)^+\boldsymbol{S} & \text{for } m > p > n. \end{cases} \tag{2.3}$$

Proof. When $n \geq m > p$ or $m > n \geq p$, the expression (2.3) follows from the fact that $\boldsymbol{S}^+ = \boldsymbol{S}^{-1}$ and

$$(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ = (\boldsymbol{X}^t)^+\boldsymbol{S}\boldsymbol{X}^+\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S}.$$

When $m > p > n$, we define the eigenvalue decomposition of $\boldsymbol{S}$ as $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$, where $\boldsymbol{H}$ is a $p \times n$ matrix such that $\boldsymbol{H}^t\boldsymbol{H} = \boldsymbol{I}_n$ and $\boldsymbol{L}$ is a full-rank diagonal matrix of order n. Since $\boldsymbol{S}^+ = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t$ and $\boldsymbol{H}^t\boldsymbol{X}^t$ is an $n \times m$ rectangular matrix of full row rank, it is observed that

$$\begin{aligned}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ &= (\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t = (\boldsymbol{H}^t\boldsymbol{X}^t)^+\boldsymbol{L}(\boldsymbol{X}\boldsymbol{H})^+\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t \\ &= \boldsymbol{X}\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H})^{-1}\boldsymbol{L}(\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t \\ &= \boldsymbol{X}\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H})^{-1}\boldsymbol{L}\boldsymbol{H}^t = \boldsymbol{X}\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t \\ &= \boldsymbol{X}\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{S}.\end{aligned}$$

Noting that $\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H})^{-1}\boldsymbol{H}^t = (\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t)^+ = (\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+)^+$, we obtain the expression for the case that $m > p > n$. □

3 A General Class of Shrinkage Estimators

Konno (1990, 1991, 1992) separately considered two cases $m > p$ and $p \geq m$, and individually defined classes of shrinkage estimators. The arguments stated in the previous section suggest that we can construct a well-defined class of shrinkage estimators unifying both cases $m > p$ and $p \geq m$.

Let $\mathcal{O}(r)$ be the group of $r \times r$ orthogonal matrices. For $r \geq q$, let $\mathcal{V}_{r,q}$ be the Stiefel manifold, namely the set of $r \times q$ matrices $\boldsymbol{M}$ such that $\boldsymbol{M}^t\boldsymbol{M} = \boldsymbol{I}_q$. It is noted that $\mathcal{O}(r) = \mathcal{V}_{r,r}$. Define $\mathbb{D}_r^+$ as the set of $r \times r$ diagonal matrices $\mathrm{diag}(d_1, \ldots, d_r)$ such that $d_1 > \cdots > d_r > 0$.

Denote $\ell = m \wedge p \wedge n$. Define the eigenvalue decomposition of $\boldsymbol{S}$ as $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$, where $\boldsymbol{H} \in \mathcal{V}_{p,n\wedge p}$ and $\boldsymbol{L} \in \mathbb{D}_{n\wedge p}^+$. Let $\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2} = \boldsymbol{R}\boldsymbol{F}^{1/2}\boldsymbol{V}^t$ be the nonsingular part of the singular value decomposition, where $\boldsymbol{R} \in \mathcal{V}_{m,\ell}$, $\boldsymbol{V} \in \mathcal{V}_{n\wedge p,\ell}$ and $\boldsymbol{F}^{1/2} = \mathrm{diag}(f_1^{1/2}, \ldots, f_\ell^{1/2}) \in \mathbb{D}_\ell^+$. It is clear that $\boldsymbol{S}^+ = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t$ and $\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t = \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t\boldsymbol{X}^t = \boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t$. Note also that $\boldsymbol{R}$ is orthogonal if $\ell = m$ and otherwise $\boldsymbol{V}$ is orthogonal.

For both the cases $m > p$ and $p \geq m$, a unified class of shrinkage estimators is defined by

$$\boldsymbol{\delta}^{SH} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+, \tag{3.1}$$

where $\boldsymbol{\Phi}(\boldsymbol{F}) = \mathrm{diag}(\phi_1(\boldsymbol{F}), \ldots, \phi_\ell(\boldsymbol{F}))$ is a diagonal matrix and the $\phi_i(\boldsymbol{F})$ are differentiable functions of $\boldsymbol{F}$. Since $(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+ = \boldsymbol{R}\boldsymbol{F}^{-1}\boldsymbol{R}^t$, the Efron-Morris type shrinkage estimator (2.2) is given by

$$\boldsymbol{\delta}^{EM} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{\Phi}^{EM}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+, \quad \boldsymbol{\Phi}^{EM}(\boldsymbol{F}) = \mathrm{diag}(c/f_1, \ldots, c/f_\ell).$$

Interestingly enough, the class (3.1) can be rewritten as in the following which is an extension of Konno's (1992) class for $m > p$.

Proposition 3.1 *Let $\boldsymbol{Q} = \boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}$. Then $\boldsymbol{Q}$ satisfies that $\boldsymbol{Q}^t\boldsymbol{S}\boldsymbol{Q} = \boldsymbol{I}_\ell$ and $\boldsymbol{Q}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{F}$, and the shrinkage estimator given in* (3.1) *is expressed as*

$$\boldsymbol{\delta}^{SH} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{Q}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{S}. \tag{3.2}$$

In the case of $\ell = n \wedge p$, namely, $m > n \wedge p$, $\boldsymbol{\delta}^{SH}$ is

$$\boldsymbol{\delta}^{SH} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{Q}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{Q}^+ = \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{Q}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{Q}^+). \tag{3.3}$$

Proof. Recall that $\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2} = \boldsymbol{R}\boldsymbol{F}^{1/2}\boldsymbol{V}^t$. It thus turns out that $\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H} = \boldsymbol{F}^{1/2}\boldsymbol{V}^t\boldsymbol{L}^{1/2}$ and $\boldsymbol{R} = \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}\boldsymbol{F}^{-1/2}$, which yields that

$$\begin{aligned}\boldsymbol{R}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ &= \boldsymbol{R}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t = \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}\boldsymbol{F}^{-1/2}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{F}^{1/2}\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\\ &= \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t.\end{aligned}$$

Then it is seen that $\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t = \boldsymbol{V}^t\boldsymbol{L}^{-1/2}\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t = \boldsymbol{Q}^t\boldsymbol{S}$. Hence for any set of (m,p,n), one gets the expression (3.2). In the case of $\ell = n \wedge p$, it is noted that $\boldsymbol{V} \in \mathcal{O}(n \wedge p)$ and $\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t = (\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V})^+ = \boldsymbol{Q}^+$, which yields the expression (3.3). □

4 A Unified Expression of the Risk Functions

We now provide a unified expression of an unbiased estimator of the risk function of estimators $\boldsymbol{\delta}^{SH}$ given in (3.1).

Theorem 4.1 *Let $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{F})$ and $\phi_i = \phi_i(\boldsymbol{F})$ for $i = 1, \ldots, \ell$. Denote by $h_{m,p}(\boldsymbol{X}|\boldsymbol{\Theta}, \boldsymbol{\Sigma})$ and $h_{n,p}(\boldsymbol{Y}|\boldsymbol{0}_{n\times p}, \boldsymbol{\Sigma})$ the probability density functions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Assume that*

(i) $E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2] < \infty$,

(ii) $\lim_{X_{ab}\to\pm\infty} \{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd} h_{m,p}(\boldsymbol{X}|\boldsymbol{\Theta}, \boldsymbol{\Sigma}) = 0$ *for $a, c = 1, \ldots, m$ and $b, d = 1, \ldots, p$,*

(iii) $\lim_{Y_{ab}\to\pm\infty} \{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd} h_{n,p}(\boldsymbol{Y}|\boldsymbol{0}_{n\times p}, \boldsymbol{\Sigma}) = 0$ *for $a, c = 1, \ldots, n$ and $b, d = 1, \ldots, p$.*

For any positive integers m, p and n, the risk difference of $\boldsymbol{\delta}^{SH}$ and $\boldsymbol{\delta}^{ML}$ is expressed as

$$\begin{aligned}&R(\boldsymbol{\delta}^{SH}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{ML}, \boldsymbol{\Theta}|\boldsymbol{\Sigma})\\ &= E\Big[\sum_{i=1}^{\ell}\Big\{af_i\phi_i^2 - 2b\phi_i - 4f_i^2\phi_i\frac{\partial\phi_i}{\partial f_i} - 4f_i\frac{\partial\phi_i}{\partial f_i}\\ &\qquad - 2\sum_{j>i}^{\ell}\frac{f_i^2\phi_i^2 - f_j^2\phi_j^2}{f_i - f_j} - 4\sum_{j>i}^{\ell}\frac{f_i\phi_i - f_j\phi_j}{f_i - f_j}\Big\}\Big],\end{aligned}$$

where $\ell = m \wedge p \wedge n$ *and*

$$\begin{aligned} a =& a_{p,m,n} = (|n-p| + 2m) \wedge (n+p) - 3, \\ b =& b_{p,m,n} = |m - n \wedge p| + 1. \end{aligned}$$

When an ordering among m, p and n is given, the corresponding specific value of (a, b) is provided. Noting that $(|n-p| + 2m) \wedge (n+p) = n+p$ for $m > n \wedge p$, we can see that specific values of (a, b) are given by

$$(a, b) = \begin{cases} (n+p-3, & m-p+1) \quad \text{for } n \geq m > p, \\ (n+p-3, & m-p+1) \quad \text{for } m > n \geq p, \\ (n+p-3, & m-n+1) \quad \text{for } m > p > n, \\ (n-p+2m-3, & p-m+1) \quad \text{for } n \geq p \geq m, \\ (p-n+2m-3, & n-m+1) \quad \text{for } p > n \geq m, \\ (n+p-3, & m-n+1) \quad \text{for } p \geq m > n. \end{cases}$$

The three cases $n \geq m > p$, $m > n \geq p$ and $n \geq p \geq m$, namely the cases satisfying $n > p$, are provided by Konno (1992).

The unified expression of the risk difference given in Theorem 4.1 can provide conditions under which specific estimators improve on the MLE $\boldsymbol{\delta}^{ML} = \boldsymbol{X}$. Two examples are given below.

Example 4.1 A modified Stein type estimator is given by

$$\boldsymbol{\delta}^{mST} = \boldsymbol{\delta}^{ST} - \frac{d}{\mathrm{tr}[\boldsymbol{X}\boldsymbol{S}^{+}\boldsymbol{X}^{t}]}\boldsymbol{R}\boldsymbol{R}^{t}\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^{+}, \tag{4.1}$$

where $\boldsymbol{\delta}^{ST} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{C}\boldsymbol{F}^{-1}\boldsymbol{R}^{t}\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^{+}$ for $\boldsymbol{C} = \mathrm{diag}(c_1, \ldots, c_\ell)$ with $c_1 \geq \cdots \geq c_\ell$. This corresponds to the form

$$\phi_i = \frac{c_i}{f_i} + \frac{d}{\sum_{j=1}^{\ell} f_j}.$$

Then, from Theorem 4.1, it follows that

$$\begin{aligned} \Delta =& R(\boldsymbol{\delta}^{mST}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{ML}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) \\ =& E\Big[\sum_{i=1}^{\ell} \frac{1}{f_i}(ac_i^2 - 2bc_i + 4c_i + 4c_i^2) \\ &+ \frac{1}{\mathrm{tr}\boldsymbol{F}}\Big\{(a - 2\ell + 2)d^2 - 2\ell bd - 2\ell(\ell-1)d + 4d + 2(a+2)d\sum_{i=1}^{\ell} c_i\Big\} \\ &+ 4d\frac{\mathrm{tr}\boldsymbol{C}\boldsymbol{F}}{(\mathrm{tr}\boldsymbol{F})^2} + 4d^2\frac{\mathrm{tr}\boldsymbol{F}^2}{(\mathrm{tr}\boldsymbol{F})^3} - 2\sum_{i=1}^{\ell}\sum_{j>i}\frac{(c_i - c_j)(c_i + c_j + 2)}{f_i - f_j} \\ &- \frac{4d}{\mathrm{tr}\boldsymbol{F}}\sum_{i=1}^{\ell}\sum_{j>i}\frac{c_i f_i - c_j f_j}{f_i - f_j}\Big], \end{aligned} \tag{4.2}$$

since $\sum_i \sum_{j>i}(f_i + f_j) = (\ell - 1)\mathrm{tr}\boldsymbol{F}$. The condition for obtaining (4.2) is

$$|m - n \wedge p| \geq 2 \qquad (\text{or, equivalently, } b \geq 3),$$

which is a sufficient condition for (i) of Theorem 4.1. It follows from Konno (1991) and Tsukuma and Kubokawa (2007) that $\mathrm{tr}\boldsymbol{F}^2 \leq (\mathrm{tr}\boldsymbol{F})^2$, $\mathrm{tr}\boldsymbol{C}\boldsymbol{F}/(\mathrm{tr}\boldsymbol{F})^2 \leq c_1/\mathrm{tr}\boldsymbol{F}$,

$$\sum_{i=1}^{\ell}\sum_{j>i}\frac{(c_i - c_j)(c_i + c_j + 2)}{f_i - f_j} \geq \sum_{i=1}^{\ell}\frac{1}{f_i}\sum_{j>i}(c_i - c_j)(c_i + c_j + 2),$$

$$\sum_{i=1}^{\ell}\sum_{j>i}\frac{c_i f_i - c_j f_j}{f_i - f_j} \geq \sum_{i=1}^{\ell}(\ell - i)c_i.$$

Thus, one gets $\Delta \leq \sum_{i=1}^{\ell} h_c(i)/f_i + h_d/\mathrm{tr}\boldsymbol{F}$, where

$$h_c(i) =(a + 4 - 2\ell + 2i)c_i^2 - 2(b - 2 + 2\ell - 2i)c_i + 2\sum_{j>i} c_j(c_j + 2),$$

$$h_d =(a - 2\ell + 6)d^2 - 2\Big\{\ell b + \ell(\ell - 1) - 2 - 2c_1 - (a + 2)\sum_{i=1}^{\ell} c_i + 2\sum_{i=1}^{\ell}(\ell - i)c_i\Big\}d.$$

Konno (1991) showed that $h_c(i) \leq 0$ for any i when c_i is

$$c_i = (b - 2 + 2\ell - 2i)/(a + 4 - 2\ell + 2i). \tag{4.3}$$

For these c_i's, it is seen that $2\sum_{i=1}^{\ell}(\ell - i)c_i = -\sum_{i=1}^{\ell}(a + 4 - 2\ell + 2i - a - 4)c_i = -\sum_{i=1}^{\ell}(b - 2 + 2\ell - 2i) + (a + 4)\sum_{i=1}^{\ell} c_i = -(b - 2 + 2\ell)\ell + \ell(\ell + 1) + (a + 4)\sum_{i=1}^{\ell} c_i$. Then, h_d is rewritten as

$$h_d = (a - 2\ell + 6)d^2 - 4\Big\{\ell - 1 + \sum_{i=2}^{\ell} c_i\Big\}d.$$

Hence, these observations imply that $\boldsymbol{\delta}^{ML}$ is improved on by the Stein (1973) type estimator $\boldsymbol{\delta}^{ST} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{C}\boldsymbol{F}^{-1}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$ for constants c_i's given in (4.3). For these c_i's, the Stein estimator $\boldsymbol{\delta}^{ST}$ is further improved on by the modified Stein type estimator

$$\boldsymbol{\delta}^{mST} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{C}\boldsymbol{F}^{-1}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ - \frac{d}{\mathrm{tr}[\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t]}\boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+,$$

if d satisfies $0 < d \leq 4\{\ell - 1 + \sum_{i=2}^{\ell} c_i\}/(a - 2\ell + 6)$. This is an extension of Tsukuma and Kubokawa (2007). □

Example 4.2 A modified Efron-Morris type estimator is given by

$$\boldsymbol{\delta}^{mEM} = \boldsymbol{\delta}^{EM} - \frac{d}{\mathrm{tr}[\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t]}\boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+, \tag{4.4}$$

where $\boldsymbol{\delta}^{EM}$ is given in (2.2). This corresponds to the form

$$\phi_i = \frac{c}{f_i} + \frac{d}{\sum_{j=1}^{\ell} f_j}.$$

Letting $c_i = c$ in (4.2) for all i, one gets

$$\begin{aligned}
&R(\boldsymbol{\delta}^{mEM}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{ML}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) \\
&= E\bigg[\{ac^2 - 2bc + 4c + 4c^2\}\sum_{i=1}^{\ell}\frac{1}{f_i} \\
&\quad + \frac{1}{\mathrm{tr}\boldsymbol{F}}\{(a - 2\ell + 2)d^2 - 2\ell bd - 2\ell(\ell-1)d + 4d + 2\ell(a+2)cd\} \\
&\quad + 4d\frac{c}{\mathrm{tr}\boldsymbol{F}} + 4d^2\frac{\mathrm{tr}\boldsymbol{F}^2}{(\mathrm{tr}\boldsymbol{F})^3} - 4\frac{cd}{\mathrm{tr}\boldsymbol{F}}\sum_{i=1}^{\ell}(\ell - i)\}\bigg],
\end{aligned}$$

which is less than or equal to

$$\begin{aligned}
E\bigg[&\{(a+4)c^2 - 2(b-2)c\}\sum_{i=1}^{\ell}\frac{1}{f_i} \\
&+ \big\{(a - 2\ell + 6)d^2 - 2d\{b\ell - 2 - (a\ell + 2\ell + 2)c + (c+1)\ell(\ell-1)\}d\big\}\frac{1}{\sum_{i=1}^{\ell} f_i}\bigg],
\end{aligned} \quad (4.5)$$

which implies that $\boldsymbol{\delta}^{mEM}$ improves on $\boldsymbol{\delta}^{ML}$ if constants c and d satisfy that $0 < c \leq 2(b-2)/(a+4)$ and

$$0 < d \leq 2\{b\ell - 2 - (a\ell + 2\ell + 2)c + (c+1)\ell(\ell-1)\}/(a - 2\ell + 6),$$

for $b > 2$ and $\{b + (c+1)(\ell - 1)\}\ell \geq 2 + (a\ell + 2\ell + 2)c$. Also from the expression (4.5), it is seen that the constant c which minimizes the first term is given by

$$c_0 = \frac{b-2}{a+4} = \frac{|m - n \wedge p| - 1}{(|n-p| + 2m) \wedge (n+p) + 1}.$$

Given $c = c_0$, the constant d which minimizes the second term in (4.5) is given by

$$d_0 = \frac{(a+b+2)(\ell-1)(\ell+2)}{(a+4)(a-2\ell+6)} = \frac{(m + n \vee p)(p \wedge m \wedge n - 1)(p \wedge m \wedge n + 2)}{\{(|n-p| + 2m) \wedge (n+p) + 1\}(|n-p| + 3)}.$$

That is, for $b \geq 3$, $\boldsymbol{\delta}^{ML}$ is improved on by the Efron-Morris type estimator

$$\boldsymbol{\delta}^{EM} = \boldsymbol{X} - c_0(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+,$$

which can be further improved on by the modified Efron-Morris type estimator

$$\boldsymbol{\delta}^{mEM} = \boldsymbol{X} - c_0(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ - \frac{d_0}{\mathrm{tr}[\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t]}\boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+,$$

for $b \geq 3$ and $\ell \geq 2$. □

5 Positive-part Estimators and Some Numerical Results

In this section we investigate risk performances of the shrinkage estimators (3.1) by simulation. Before that, we prove that the shrinkage estimators are dominated by the corresponding positive-part shrinkage estimators.

Since $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$ and $\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2} = \boldsymbol{R}\boldsymbol{F}^{1/2}\boldsymbol{V}^t$ where $\boldsymbol{R} \in \mathcal{V}_{m,\ell}$, it is seen that

$$(\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t)\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ = (\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t)\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t = \boldsymbol{0}_{m\times p},$$

which is used to rewrite the estimator (3.1) as

$$\begin{aligned}\boldsymbol{\delta}^{SH} &= \boldsymbol{X} - \boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ + \boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ - \boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ + \boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ - \boldsymbol{R}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+ \\ &= \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+) + \boldsymbol{R}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+,\end{aligned}$$

where $\boldsymbol{\Psi}(\boldsymbol{F}) = \mathrm{diag}(\psi_1(\boldsymbol{F}), \ldots, \psi_\ell(\boldsymbol{F})) = \boldsymbol{I}_\ell - \boldsymbol{\Phi}(\boldsymbol{F})$. Then, we define the positive-part shrinkage estimator

$$\boldsymbol{\delta}_+^{SH} = \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+) + \boldsymbol{R}\boldsymbol{\Psi}_+(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+, \tag{5.1}$$

where $\boldsymbol{\Psi}_+(\boldsymbol{F}) = \mathrm{diag}(\psi_1^+(\boldsymbol{F}), \ldots, \psi_\ell^+(\boldsymbol{F}))$ for $\psi_i^+(\boldsymbol{F}) = \max\{0, \psi_i(\boldsymbol{F})\}$.

When $m = 1$ and $p > n$, $\boldsymbol{\delta}_+^{SH}$ was suggested by Chételat and Wells (2012), who showed by simulation that $\boldsymbol{\delta}_+^{SH}$ outperforms $\boldsymbol{\delta}^{SH}$. For analytical dominance results between $\boldsymbol{\delta}_+^{SH}$ and $\boldsymbol{\delta}^{SH}$, see Baranchik (1970) for $m = 1$ and $n \geq p$ and Tsukuma (2010) for $m > 1$ and $n \geq p$. We prove analytically this kind of dominance results in more general cases for any positive numbers m, p and n. The proof of the following theorem is given in Section 6.

Theorem 5.1 *Assume that the risk of $\boldsymbol{\delta}^{SH}$ is finite and* $\Pr(\psi_i(\boldsymbol{F}) < 0) > 0$ *for some i. Then $\boldsymbol{\delta}_+^{SH}$ dominates $\boldsymbol{\delta}^{SH}$ relative to the loss* (1.2) *regardless of an order relation among m, p and n.*

For example, the Efron-Morris estimator $\boldsymbol{\delta}^{EM}$ is dominated by $\boldsymbol{\delta}_+^{EM} = \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+) + \boldsymbol{R}\boldsymbol{\Psi}_+^{EM}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$, where the i-th diagonal element of $\boldsymbol{\Psi}_+^{EM}(\boldsymbol{F})$ is $\max[0, 1 - (b-2)/\{(a+4)f_i\}]$. Also, Theorem 5.1 can be applied to $\boldsymbol{\delta}^{mEM}$, $\boldsymbol{\delta}^{ST}$ and $\boldsymbol{\delta}^{mST}$ given in Section 4.

We now investigate how positive-part shrinkage estimators reduce risks of shrinkage estimators through Monte Carlo simulations. The risks of estimators were estimated by average of losses based on 10,000 independent replications of $\boldsymbol{X}$ and $\boldsymbol{Y}$ in the model (1.1). For the mean matrix $\boldsymbol{\Theta} = (\theta_{ij})$, we considered the following two cases: (A) $\boldsymbol{\Theta} = \boldsymbol{0}_{m\times p}$ and (B) $\theta_{ij} = 2\sin(i^2 + j)$ for $i = 1, \ldots, m$ and $j = 1, \ldots, p$. For the covariance matrix $\boldsymbol{\Sigma}$, we supposed (a) $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ or (b) $\boldsymbol{\Sigma} = \mathrm{diag}(1, 2^{-1}, \ldots, p^{-1})$. Also, m, p and n were taken as $(m,p) = (20, 10)$ and $(10, 20)$ and $n = 50, 15$ and 5. Then the risk of $\boldsymbol{\delta}^{ML} = \boldsymbol{X}$ is given by $mp = 200$.

In the simulations, we examined the following shrinkage estimators:

Table 1: Average Losses of Shrinkage Estimators and Their Positive-part Estimators ($R(\boldsymbol{X}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = mp = 200$)

$\boldsymbol{\Theta}$	$\boldsymbol{\Sigma}$	(m,p)	n	$\boldsymbol{\delta}^{EM}$	$\boldsymbol{\delta}^{EM}_+$	$\boldsymbol{\delta}^{mEM}$	$\boldsymbol{\delta}^{mEM}_+$	$\boldsymbol{\delta}^{ST}$	$\boldsymbol{\delta}^{ST}_+$	$\boldsymbol{\delta}^{mST}$	$\boldsymbol{\delta}^{mST}_+$
(A)	(a)	$(20,10)$	50	126.2	106.9	49.8	11.7	57.3	28.3	52.2	17.4
			15	147.9	135.5	64.1	41.8	77.1	59.8	63.7	42.4
			5	156.5	149.6	134.3	121.0	136.5	128.6	131.0	120.5
		$(10,20)$	50	129.2	111.1	49.4	13.2	58.4	31.2	52.2	19.3
			15	177.3	169.2	94.3	81.6	109.9	97.0	97.4	81.7
			5	184.7	180.0	164.4	155.8	167.4	161.4	164.1	156.3
(A)	(b)	$(20,10)$	50	126.2	106.9	49.8	11.7	57.3	28.3	52.2	17.4
			15	147.9	135.5	64.1	41.8	77.1	59.8	63.7	42.4
			5	162.7	157.0	144.4	133.8	146.2	139.8	141.7	133.4
		$(10,20)$	50	129.2	111.1	49.4	13.2	58.4	31.2	52.2	19.3
			15	178.9	171.5	103.2	91.7	117.0	105.4	105.7	91.5
			5	189.4	186.4	176.5	171.6	178.4	174.8	176.3	171.9
(B)	(a)	$(20,10)$	50	139.8	125.5	114.6	96.4	97.6	77.5	96.2	74.8
			15	157.5	148.2	129.5	118.0	115.7	103.5	111.9	98.9
			5	170.6	166.8	162.9	158.0	162.5	158.5	160.9	156.4
		$(10,20)$	50	142.3	128.7	116.0	98.7	99.1	80.3	97.4	77.3
			15	181.6	175.3	153.7	146.4	140.6	131.5	137.0	127.3
			5	189.4	186.8	182.4	179.1	181.9	178.9	180.9	177.6
(B)	(b)	$(20,10)$	50	140.8	126.8	134.5	119.6	99.9	80.5	99.6	79.8
			15	158.1	149.0	151.1	141.6	119.3	107.6	118.5	106.5
			5	177.3	174.8	175.0	172.3	172.6	170.0	172.2	169.5
		$(10,20)$	50	143.2	130.1	139.7	126.1	101.8	83.7	101.6	83.3
			15	183.1	177.7	179.0	173.4	148.8	141.0	148.3	140.5
			5	193.6	192.3	192.3	190.9	190.3	188.9	190.1	188.7

(1) $\boldsymbol{\delta}^{EM} = \boldsymbol{X} - c_0 \boldsymbol{R}\boldsymbol{F}^{-1}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$, $c_0 = \dfrac{b-2}{a+4}$;

(2) $\boldsymbol{\delta}^{mEM} = \boldsymbol{\delta}^{EM} - d_0(\mathrm{tr}\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^{-1}\boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$, $d_0 = \dfrac{(a+b+2)(\ell-1)(\ell+2)}{(a+4)(a-2\ell+6)}$;

(3) $\boldsymbol{\delta}^{ST} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{C}\boldsymbol{F}^{-1}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$, $\boldsymbol{C} = \mathrm{diag}(c_1, \ldots, c_\ell)$ with $c_i = \dfrac{b-2+2\ell-2i}{a+4-2\ell+2i}$;

(4) $\boldsymbol{\delta}^{mST} = \boldsymbol{\delta}^{ST} - d_1(\mathrm{tr}\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^{-1}\boldsymbol{R}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$, $d_1 = \dfrac{2\{\ell-1+\sum_{i=2}^{\ell} c_i\}}{a-2\ell+6}$.

The corresponding positive-part estimators are denoted by $\boldsymbol{\delta}^{EM}_+$, $\boldsymbol{\delta}^{mEM}_+$, $\boldsymbol{\delta}^{ST}_+$ and $\boldsymbol{\delta}^{mST}_+$, respectively.

It is noted that in the cases such that $n \geq p$, the estimation problem, shrinkage estimators $\boldsymbol{\delta}^{SH}$ and positive-part estimators $\boldsymbol{\delta}_+^{SH}$ are invariant under the transformations

$$\boldsymbol{X} \to \boldsymbol{O}\boldsymbol{X}\boldsymbol{P}, \quad \boldsymbol{\Theta} \to \boldsymbol{O}\boldsymbol{\Theta}\boldsymbol{P}, \quad \boldsymbol{S} \to \boldsymbol{P}^t\boldsymbol{S}\boldsymbol{P}, \quad \boldsymbol{\Sigma} \to \boldsymbol{P}^t\boldsymbol{\Sigma}\boldsymbol{P},$$

where $\boldsymbol{O} \in \mathcal{O}(m)$ and $\boldsymbol{P}$ is a $p \times p$ nonsingular matrix. Then the risk functions of $\boldsymbol{\delta}^{SH}$ and $\boldsymbol{\delta}_+^{SH}$ are functions of eigenvalues of $\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Theta}^t$. However, in the cases such that $p > n$, $\boldsymbol{\delta}^{SH}$ and $\boldsymbol{\delta}_+^{SH}$ are not invariant under the above transformations because $(\boldsymbol{P}^t\boldsymbol{S}\boldsymbol{P})^+ \neq \boldsymbol{P}^{-1}\boldsymbol{S}^+(\boldsymbol{P}^t)^{-1}$.

Our findings of the simulations are summarized in Table 1. When $\boldsymbol{\Theta} = \boldsymbol{0}_{m\times p}$ with $n = 50$, the risk improvement of positive-part estimator over the corresponding shrinkage estimator is very substantial. Through all cases, $\boldsymbol{\delta}_+^{mST}$ provide large savings in risk. The simulation results also suggest that, as n is small, shrinkage and positive-part estimators are less effective.

6 Proofs

6.1 Proof of Theorem 4.1

For an $m \times p$ rectangular matrix $\boldsymbol{X} = (X_{ab})$, define the $m \times p$ rectangular matrix of differential operators with respect to $\boldsymbol{X}$ as $\nabla_X = \big(\mathrm{d}_{ab}^X\big)$, where $\mathrm{d}_{ab}^X = \partial/\partial X_{ab}$. Similarly, denote by $\nabla_Y = \big(\mathrm{d}_{ab}^Y\big)$ the $n \times p$ rectangular matrix of differential operators with respect to an $n \times p$ rectangular matrix $\boldsymbol{Y} = (Y_{ab})$.

A key tool for deriving the unbiased estimator of the risk function is the Stein identity, which is given in the following lemma. For details, see Kubokawa and Srivastava (2001).

Lemma 6.1 *Let $\boldsymbol{X}$ be defined as in the model* (1.1)*. Let $\boldsymbol{\Theta} = (\theta_{ab})$ and denote by $h_{m,p}(\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{\Sigma})$ the probability density function (p.d.f.) of $\boldsymbol{X}$. Let $\boldsymbol{G} = (G_{cd})$ be an $m \times p$ matrix such that all the elements G_{cd} are absolutely continuous functions of $\boldsymbol{X}$ and satisfy $E[|(X_{ab} - \theta_{ab})G_{cd}|] < \infty$ and $\lim_{X_{ab}\to\pm\infty} G_{cd}h_{m,p}(\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{\Sigma}) = 0$ for $a, c = 1, \ldots, m$ and $b, d = 1, \ldots, p$. It then follows that*

$$E[\mathrm{tr}(\boldsymbol{X} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^t] = E[\mathrm{tr}\nabla_X\boldsymbol{G}^t].$$

Recall that $\boldsymbol{S} = \boldsymbol{Y}^t\boldsymbol{Y}$, where $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^t$ with $\boldsymbol{Y}_i \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Sigma})$. Konno (2009) used Lemma 6.1 to obtain the identity

$$E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{G}^t] = E[\mathrm{tr}\nabla_Y^t\boldsymbol{Y}\boldsymbol{G}^t], \tag{6.1}$$

where $\boldsymbol{G}$ is a $p \times p$ matrix-valued function of $\boldsymbol{Y}$. This identity is also useful for evaluating the risk in high dimensions.

Next, we provide calculus formulas for a $p \times p$ symmetric matrix $\boldsymbol{S} = (S_{ij}) = \boldsymbol{Y}^t\boldsymbol{Y}$ and its Moore-Penrose inverse $\boldsymbol{S}^+ = (S_{ij}^+)$.

Lemma 6.2 *Denote the Kronecker delta by δ_{ij}, namely $\delta_{ij}=1$ for $i=j$ and $\delta_{ij}=0$ for $i\neq j$. For $a=1,\ldots,n$ and $b=1,\ldots,p$, we have*

(i) $\mathrm{d}^Y_{ab}S_{cd}=\delta_{bc}Y_{ad}+\delta_{bd}Y_{ac}$ *for $c,d=1,\ldots,p$,*

(ii) $$\mathrm{d}^Y_{ab}S^+_{cd}=-S^+_{bc}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ad}-S^+_{bd}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ac}+\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{bc}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\}_{ad} +\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{bd}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\}_{ac} \quad \text{for } c,d=1,\ldots,p,$$

(iii) $$\mathrm{d}^Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{cd}=\{\boldsymbol{I}_n-\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ac}S^+_{bd}+\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ac}\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{bd} -\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ad}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{cb} \quad \text{for } c=1,\ldots,n \text{ and } d=1,\ldots,p.$$

Proof. For the proof of (i), see Chételat and Wells (2012, (i) of Proposition 1).

Since the differential of $\boldsymbol{S}^+$ is given by

$$\mathrm{d}\boldsymbol{S}^+=-\boldsymbol{S}^+(\mathrm{d}\boldsymbol{S})\boldsymbol{S}^++(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)(\mathrm{d}\boldsymbol{S})\boldsymbol{S}^+\boldsymbol{S}^++\boldsymbol{S}^+\boldsymbol{S}^+(\mathrm{d}\boldsymbol{S})(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+),$$

it is observed that from (i)

$$\begin{aligned}\mathrm{d}^Y_{ab}S^+_{cd}=\{\mathrm{d}^Y_{ab}\boldsymbol{S}^+\}_{cd}&=\sum_{i=1}^p\sum_{j=1}^p\Big[-S^+_{ci}(\mathrm{d}^Y_{ab}S_{ij})S^+_{jd}+\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{ci}(\mathrm{d}^Y_{ab}S_{ij})\{\boldsymbol{S}^+\boldsymbol{S}^+\}_{jd}\\&\qquad+\{\boldsymbol{S}^+\boldsymbol{S}^+\}_{ci}(\mathrm{d}^Y_{ab}S_{ij})\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{jd}\Big]\\&=-S^+_{bc}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ad}-\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ac}S^+_{bd}\\&\quad+\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{bc}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\}_{ad}+\{\boldsymbol{Y}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\}_{ac}\{\boldsymbol{S}^+\boldsymbol{S}^+\}_{bd}\\&\quad+\{\boldsymbol{S}^+\boldsymbol{S}^+\}_{bc}\{\boldsymbol{Y}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\}_{ad}+\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\}_{ac}\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{bd}.\end{aligned}$$

Noting that $\boldsymbol{Y}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)=\boldsymbol{0}_{n\times p}$, we get (ii).

The product rule is used to obtain

$$\mathrm{d}^Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{cd}=\sum_{i=1}^p\{(\mathrm{d}^Y_{ab}Y_{ci})S^+_{id}+Y_{ci}\mathrm{d}^Y_{ab}S^+_{id}\}.$$

Using (ii) and summing up with respect to i yields (iii). □

Recall that $\boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t$ denotes the eigenvalue decomposition of $\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t$, where $\boldsymbol{R}=(R_{ij})\in\mathcal{V}_{m,\ell}$ and $\boldsymbol{F}=\mathrm{diag}(f_1,\ldots,f_\ell)\in\mathcal{D}^+_\ell$. The following lemma shows partial derivatives of $\boldsymbol{F}$ and $\boldsymbol{R}$ with respect to ∇_X and ∇_Y.

Lemma 6.3 *For $i=1,\ldots,\ell$, $k=1,\ldots,m$, $a=1,\ldots,m$ and $b=1,\ldots,p$, we have*

(i) $\mathrm{d}^X_{ab}f_i=A^{ii}_{1\cdot ab}$,

(ii) $$\mathrm{d}^X_{ab}R_{ki}=\sum_{\substack{j=1\\j\neq i}}^{\ell}\frac{R_{kj}A^{ij}_{1\cdot ab}}{f_i-f_j}+f_i^{-1}\{\boldsymbol{I}_m-\boldsymbol{R}\boldsymbol{R}^t\}_{ak}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib},$$

where $A^{ij}_{1\cdot ab} = R_{aj}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib} + R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{jb}$. *For* $i = 1,\dots,\ell$, $k = 1,\dots,m$, $a = 1,\dots,n$ *and* $b = 1,\dots,p$, *we have*

(iii) $\mathrm{d}^Y_{ab} f_i = B^{ii}_{1\cdot ab}$,

(iv) $\displaystyle \mathrm{d}^Y_{ab} R_{ki} = \sum_{\substack{j=1\\ j\neq i}}^{\ell} \frac{R_{kj}B^{ij}_{1\cdot ab}}{f_i - f_j} + f_i^{-1}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{(\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t)\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{kb}$,

where

$$\begin{aligned} B^{ij}_{1\cdot ab} &= -\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{jb} - \{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ja}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib} \\ &\quad + \{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{jb} \\ &\quad + \{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ja}\{\boldsymbol{R}^t\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{ib}. \end{aligned}$$

Proof. Take $\boldsymbol{R}_0 \in \mathcal{V}_{m,m-\ell}$ such that $\boldsymbol{R}_0^t\boldsymbol{R} = \boldsymbol{0}_{(m-\ell)\times\ell}$. Define $\boldsymbol{U} = (U_{ij}) = [\boldsymbol{R}, \boldsymbol{R}_0] \in \mathcal{O}_m$. Denote $\boldsymbol{F}_0 = \mathrm{diag}(f_1,\dots,f_\ell,0,\dots,0)$, where $\boldsymbol{F}_0$ is of order m. It is clear that $\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t = \boldsymbol{U}\boldsymbol{F}_0\boldsymbol{U}^t$. Since the differential of $\boldsymbol{U}^t\boldsymbol{U} = \boldsymbol{I}_m$ is given by $(\mathrm{d}\boldsymbol{U}^t)\boldsymbol{U} + \boldsymbol{U}^t(\mathrm{d}\boldsymbol{U}) = \boldsymbol{0}_{m\times m}$, the $m \times m$ matrix $\boldsymbol{U}^t(\mathrm{d}\boldsymbol{U})$ is skew-symmetric, namely the (j,i)-th element is written as

$$\{\boldsymbol{U}^t(\mathrm{d}\boldsymbol{U})\}_{ji} = \begin{cases} 0 & \text{for } j = i, \\ -\{(\mathrm{d}\boldsymbol{U}^t)\boldsymbol{U}\}_{ji} & \text{for } j \neq i. \end{cases}$$

The differential of $\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t = \boldsymbol{U}\boldsymbol{F}_0\boldsymbol{U}^t$ is given by

$$\mathrm{d}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t) = (\mathrm{d}\boldsymbol{U})\boldsymbol{F}_0\boldsymbol{U}^t + \boldsymbol{U}(\mathrm{d}\boldsymbol{F}_0)\boldsymbol{U}^t + \boldsymbol{U}\boldsymbol{F}_0(\mathrm{d}\boldsymbol{U}^t),$$

which yields

$$\begin{aligned} \boldsymbol{U}^t[\mathrm{d}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U} &= \boldsymbol{U}^t(\mathrm{d}\boldsymbol{U})\boldsymbol{F}_0 + \mathrm{d}\boldsymbol{F}_0 + \boldsymbol{F}_0(\mathrm{d}\boldsymbol{U}^t)\boldsymbol{U} \\ &= \boldsymbol{U}^t(\mathrm{d}\boldsymbol{U})\boldsymbol{F}_0 + \mathrm{d}\boldsymbol{F}_0 - \boldsymbol{F}_0\boldsymbol{U}^t(\mathrm{d}\boldsymbol{U}). \end{aligned}$$

It is thus seen that

$$\mathrm{d}f_i = \{\boldsymbol{U}^t[\mathrm{d}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ii} \quad \text{for } i = 1,\dots,\ell$$

and

$$\{\boldsymbol{U}^t(\mathrm{d}\boldsymbol{U})\}_{ji} = \begin{cases} \dfrac{\{\boldsymbol{U}^t[\mathrm{d}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji}}{f_i - f_j} & \text{for } j = 1,\dots,\ell \text{ and } i = 1,\dots,\ell \text{ with } j \neq i, \\ \dfrac{\{\boldsymbol{U}^t[\mathrm{d}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji}}{f_i} & \text{for } j = \ell+1,\dots,m \text{ and } i = 1,\dots,\ell. \end{cases}$$

Noting that $\mathrm{d}^X_{ab}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t) = (\mathrm{d}^X_{ab}\boldsymbol{X})\boldsymbol{S}^+\boldsymbol{X}^t + \boldsymbol{X}\boldsymbol{S}^+(\mathrm{d}^X_{ab}\boldsymbol{X}^t)$ and $\mathrm{d}^X_{ab}X_{cd} = \delta_{ac}\delta_{bd}$, we observe that

$$\begin{aligned} &\{\boldsymbol{U}^t[\mathrm{d}^X_{ab}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji} \\ &= \sum_{c=1}^{m}\sum_{d=1}^{p} U_{cj}(\mathrm{d}^X_{ab}X_{cd})\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{U}\}_{di} + \sum_{c=1}^{p}\sum_{d=1}^{m}\{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\}_{jc}(\mathrm{d}^X_{ab}X_{dc})U_{di} \\ &= U_{aj}\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{U}\}_{bi} + \{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\}_{jb}U_{ai}. \end{aligned}$$

Since $\{\boldsymbol{U}^t[\mathrm{d}_{ab}^X(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji} = A_{1\cdot ab}^{ij}$ for $i,j=1,\dots,\ell$, it follows that

$$\mathrm{d}_{ab}^X f_i = \{\boldsymbol{U}^t[\mathrm{d}_{ab}^X(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ii} = A_{1\cdot ab}^{ii},$$

which gives (i). It is also observed that for $k=1,\dots,m$ and $i=1,\dots,\ell$

$$\begin{aligned}\mathrm{d}_{ab}^X R_{ki} &= \{\mathrm{d}_{ab}^X\boldsymbol{U}\}_{ki} = \{\boldsymbol{U}\boldsymbol{U}^t(\mathrm{d}_{ab}^X\boldsymbol{U})\}_{ki}\\ &= \sum_{\substack{j=1\\ j\neq i}}^{\ell} U_{kj}\{\boldsymbol{U}^t(\mathrm{d}_{ab}^X\boldsymbol{U})\}_{ji} + \sum_{j=\ell+1}^{m} U_{kj}\{\boldsymbol{U}^t(\mathrm{d}_{ab}^X\boldsymbol{U})\}_{ji}\\ &= \sum_{\substack{j=1\\ j\neq i}}^{\ell} \frac{R_{kj}A_{1\cdot ab}^{ij}}{f_i-f_j} + \sum_{j=\ell+1}^{m} \frac{U_{kj}\{\boldsymbol{U}^t[\mathrm{d}_{ab}^X(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji}}{f_i}. \end{aligned} \tag{6.2}$$

Here it is seen that

$$\begin{aligned}\sum_{j=\ell+1}^{m} U_{kj}\{\boldsymbol{U}^t[\mathrm{d}_{ab}^X(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji} &= \{\boldsymbol{R}_0\boldsymbol{R}_0^t\}_{ak}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib} + R_{ai}\{\boldsymbol{R}_0\boldsymbol{R}_0^t\boldsymbol{X}\boldsymbol{S}^+\}_{kb}\\ &= \{\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t\}_{ak}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib} \end{aligned} \tag{6.3}$$

because $\boldsymbol{R}\boldsymbol{R}^t + \boldsymbol{R}_0\boldsymbol{R}_0^t = \boldsymbol{I}_m$ and $\boldsymbol{R}_0^t\boldsymbol{X}\boldsymbol{S}^+ = \boldsymbol{0}_{(m-\ell)\times p}$. Substituting (6.3) into (6.2), we obtain (ii).

Since $\{\boldsymbol{U}^t[\mathrm{d}_{ab}^Y(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji} = \{\boldsymbol{U}^t\boldsymbol{X}[\mathrm{d}_{ab}^Y\boldsymbol{S}^+]\boldsymbol{X}^t\boldsymbol{U}\}_{ji}$, it is observed that from (ii) of Lemma 6.2

$$\begin{aligned}&\{\boldsymbol{U}^t[\mathrm{d}_{ab}^Y(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji}\\ &= -\{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\}_{jb}\{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia} - \{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ja}\{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib}\\ &\quad + \{\boldsymbol{U}^t\boldsymbol{X}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\}_{jb}\{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia} + \{\boldsymbol{U}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ja}\{\boldsymbol{U}^t\boldsymbol{X}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\}_{ib}.\end{aligned}$$

It is noted that $\{\boldsymbol{U}^t[\mathrm{d}_{ab}^Y(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji} = B_{1\cdot ab}^{ij}$ for $i,j=1,\dots,\ell$ and

$$\sum_{j=\ell+1}^{m} U_{kj}\{\boldsymbol{U}^t[\mathrm{d}_{ab}^Y(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)]\boldsymbol{U}\}_{ji} = \{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{(\boldsymbol{I}_m-\boldsymbol{R}\boldsymbol{R}^t)\boldsymbol{X}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\}_{kb}.$$

Hence using the same arguments as in the proofs of (i) and (ii) yields (iii) and (iv). □

Lemma 6.4 *Let* $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{F}) = \mathrm{diag}(\phi_1,\dots,\phi_\ell)$ *be a diagonal matrix of order* ℓ*, where the* ϕ_i *are differentiable functions of* $\boldsymbol{F}$*. For* $a=1,\dots,m$*,* $b=1,\dots,p$ *and* $i,j=1,\dots,\ell$*, denote* $A_{2\cdot ab}^{ij} = R_{aj}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib} + R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{jb}$*. Let* $A_{1\cdot ab}^{ij}$*'s be defined as in Lemma 6.3. Then we have*

$$\mathrm{d}_{ab}^X\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab} = \{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\}_{aa}\{\boldsymbol{S}\boldsymbol{S}^+\}_{bb} + D_{ab}^{X\cdot 1}(\boldsymbol{\Phi}) + D_{ab}^{X\cdot 2}(\boldsymbol{\Phi}),$$

where

$$D_{ab}^{X\cdot 1}(\mathbf{\Phi}) = \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} A_{1\cdot ab}^{jj}A_{2\cdot ab}^{ii}\frac{\partial \phi_i}{\partial f_j},$$

$$D_{ab}^{X\cdot 2}(\mathbf{\Phi}) = \{\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t\}_{aa}\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{F}^{-1}\mathbf{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb} + \sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{\phi_i}{f_i - f_j}A_{1\cdot ab}^{ij}A_{2\cdot ab}^{ij}.$$

Proof. Since $\{\boldsymbol{R}\mathbf{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab} = \sum_{i=1}^{\ell}\sum_{k=1}^{m} R_{ai}\phi_i R_{ki}\{\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{kb}$, it is seen that

$$\begin{aligned}\mathrm{d}_{ab}^X\{\boldsymbol{R}\mathbf{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab} &= \sum_{k=1}^{m}\{\boldsymbol{R}\mathbf{\Phi}\boldsymbol{R}^t\}_{ak}\mathrm{d}_{ab}^X\{\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{kb} + D_1 + D_2\\ &= \{\boldsymbol{R}\mathbf{\Phi}\boldsymbol{R}^t\}_{aa}\{\boldsymbol{S}\boldsymbol{S}^+\}_{bb} + D_1 + D_2,\end{aligned}$$

where

$$D_1 = \sum_{i=1}^{\ell} R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}\mathrm{d}_{ab}^X\phi_i,$$

$$D_2 = \sum_{i=1}^{\ell}\sum_{k=1}^{m}\phi_i\{\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{kb}(R_{ki}\mathrm{d}_{ab}^X R_{ai} + R_{ai}\mathrm{d}_{ab}^X R_{ki}).$$

It will be shown that $D_1 = D_{ab}^{X\cdot 1}(\mathbf{\Phi})$ and $D_2 = D_{ab}^{X\cdot 2}(\mathbf{\Phi})$.

Applying the chain rule and (i) of Lemma 6.3 to D_1 gives that

$$D_1 = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}(\mathrm{d}_{ab}^X f_j)\frac{\partial \phi_i}{\partial f_j} = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}A_{1\cdot ab}^{jj}\frac{\partial \phi_i}{\partial f_j}.$$

It follows that $R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib} = (1/2)A_{2\cdot ab}^{ii}$, which implies that $D_1 = D_{ab}^{X\cdot 1}(\mathbf{\Phi})$.

It is noted that from (ii) of Lemma 6.3

$$\begin{aligned}\sum_{i=1}^{\ell}\sum_{k=1}^{m}\phi_i\{\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{kb}R_{ki}\mathrm{d}_{ab}^X R_{ai} &= \sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{\phi_i}{f_i - f_j}A_{1\cdot ab}^{ij}R_{aj}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}\\ &\quad + \{\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t\}_{aa}\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{F}^{-1}\mathbf{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb}\end{aligned}$$

and

$$\sum_{i=1}^{\ell}\sum_{k=1}^{m}\phi_i\{\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{kb}R_{ai}\mathrm{d}_{ab}^X R_{ki} = \sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{\phi_i}{f_i - f_j}A_{1\cdot ab}^{ij}R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{jb},$$

which yields that

$$\begin{aligned}D_2 = \sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{\phi_i}{f_i - f_j}A_{1\cdot ab}^{ij}(R_{aj}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib} + R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{jb})\\ + \{\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t\}_{aa}\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{F}^{-1}\mathbf{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb}.\end{aligned}$$

Since $R_{aj}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}+R_{ai}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{jb}=A^{ij}_{2\cdot ab}$, we can see that $D_2=D^{X\cdot 2}_{ab}(\boldsymbol{\Phi})$. Thus the proof is complete. □

Lemma 6.5 *Let $B^{ij}_{1\cdot ab}$'s and $\boldsymbol{\Phi}$ be defined as in Lemmas 6.3 and 6.4, respectively. For $a=1,\ldots,n$, $b=1,\ldots,p$ and $i,j=1,\ldots,\ell$, denote $B^{ij}_{2\cdot ab}=\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{jb}+\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ja}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}$. Then we have*

$$\mathrm{d}^Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab}=D^{Y\cdot 1}_{ab}(\boldsymbol{\Phi})+D^{Y\cdot 2}_{ab}(\boldsymbol{\Phi})+D^{Y\cdot 3}_{ab}(\boldsymbol{\Phi}),$$

where

$$\begin{aligned}
D^{Y\cdot 1}_{ab}(\boldsymbol{\Phi})=&\{\boldsymbol{I}_n-\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb}\\
&+\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}\{(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb}\\
&-\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ab}\\
&+\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\}_{ab}\\
&+\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}\{\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+\}_{bb},\\
D^{Y\cdot 2}_{ab}(\boldsymbol{\Phi})=&\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}B^{jj}_{1\cdot ab}B^{ii}_{2\cdot ab}\phi_i\frac{\partial\phi_i}{\partial f_j},\\
D^{Y\cdot 3}_{ab}(\boldsymbol{\Phi})=&\sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{\phi_i^2}{f_i-f_j}B^{ij}_{1\cdot ab}B^{ij}_{2\cdot ab}.
\end{aligned}$$

Proof. It is observed that

$$\begin{aligned}
\mathrm{d}^Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab}&=\mathrm{d}^Y_{ab}\sum_{i=1}^{\ell}\sum_{c=1}^{n}\phi_i^2\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ic}Y_{cb}\\
&=D_1+D_2+\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}, \qquad (6.4)
\end{aligned}$$

where

$$\begin{aligned}
D_1=&\sum_{i=1}^{\ell}\phi_i^2\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}\mathrm{d}^Y_{ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\\
&+\sum_{i=1}^{\ell}\sum_{c=1}^{n}\phi_i^2\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}Y_{cb}\mathrm{d}^Y_{ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ic},\\
D_2=&\sum_{i=1}^{\ell}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}\mathrm{d}^Y_{ab}\phi_i^2.
\end{aligned}$$

Using the chain rule and (iii) of Lemma 6.3, we express D_2 of (6.4) as

$$D_2=2\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib}B^{jj}_{1\cdot ab}\phi_i\frac{\partial\phi_i}{\partial f_j}.$$

Since $\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ib} = (1/2)B^{ii}_{2\cdot ab}$, we get

$$D_2 = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} B^{jj}_{1\cdot ab}B^{ii}_{2\cdot ab}\phi_i\frac{\partial\phi_i}{\partial f_j} = D^{Y\cdot 2}_{ab}(\boldsymbol{\Phi}). \tag{6.5}$$

We next evaluate $\mathrm{d}^Y_{ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ic}$ for $i = 1,\ldots,\ell$ and $c = 1,\ldots,n$. Using the product rule gives that

$$\mathrm{d}^Y_{ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ic} = \sum_{k=1}^{m}\{\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{kc}\mathrm{d}^Y_{ab}R_{ki} + \sum_{k=1}^{p}\{\boldsymbol{R}^t\boldsymbol{X}\}_{ik}\mathrm{d}^Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ck}. \tag{6.6}$$

From (iv) of Lemma 6.3, the first term of the r.h.s. in (6.6) is written as

$$\sum_{k=1}^{m}\{\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{kc}\mathrm{d}^Y_{ab}R_{ki} = \sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{B^{ij}_{1\cdot ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{jc}}{f_i - f_j}. \tag{6.7}$$

because $\sum_{k=1}^m\{\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ka}\{(\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{R}^t)\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{kb} = 0$. Applying (iii) of Lemma 6.2 to the second term of the r.h.s. in (6.6) gives that

$$\begin{aligned}\sum_{k=1}^{p}\{\boldsymbol{R}^t\boldsymbol{X}\}_{ik}\mathrm{d}^Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ck} &= \{\boldsymbol{I}_n - \boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ac}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib} - \{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{cb}\\ &\quad + \{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ac}\{\boldsymbol{R}^t\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{ib}.\end{aligned} \tag{6.8}$$

Combining (6.6), (6.7) and (6.8), we obtain

$$\begin{aligned}\mathrm{d}^Y_{ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ic} &= \sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{B^{ij}_{1\cdot ab}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{jc}}{f_i - f_j} + \{\boldsymbol{I}_n - \boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ac}\{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\}_{ib}\\ &\quad - \{\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ia}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{cb} + \{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{ac}\{\boldsymbol{R}^t\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{ib}.\end{aligned} \tag{6.9}$$

Applying (6.9) to D_1 of (6.4) implies that

$$\begin{aligned}D_1 &= \sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{\phi_i^2}{f_i - f_j}B^{ij}_{1\cdot ab}B^{ij}_{2\cdot ab} + \{\boldsymbol{I}_n - \boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}\{\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb}\\ &\quad + \{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}\{(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{bb}\\ &\quad - \{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\}_{ab} - \{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{Y}^t\}_{aa}\{\boldsymbol{S}\boldsymbol{S}^+\}_{bb}\\ &\quad + \{\boldsymbol{Y}\boldsymbol{S}^+\}_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\}_{ab}.\end{aligned}$$

The first term of the above r.h.s. is equal to $D^{Y\cdot 3}_{ab}(\boldsymbol{\Phi})$. Hence the sum of D_1 and the third term of the last r.h.s. in (6.4) is $D^{Y\cdot 1}_{ab}(\boldsymbol{\Phi}) + D^{Y\cdot 3}_{ab}(\boldsymbol{\Phi})$. Combining this result and (6.5) completes the proof. □

Proof of Theorem 4.1. Abbreviate $\boldsymbol{\Phi}(\boldsymbol{F})$ to $\boldsymbol{\Phi}$. The risk of $\boldsymbol{\delta}^{SH}$ is expanded as

$$\begin{aligned} R(\boldsymbol{\delta}^{SH}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) &= R(\boldsymbol{X}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) + E[\mathrm{tr}\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t \\ &\qquad - 2\mathrm{tr}(\boldsymbol{X}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t] \\ &= R(\boldsymbol{X}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) + E_2 - 2E_1, \end{aligned} \tag{6.10}$$

where

$$\begin{aligned} E_1 &= E[\mathrm{tr}(\boldsymbol{X}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t], \\ E_2 &= E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+]. \end{aligned}$$

Using Lemma 6.1, we can express E_1 as

$$E_1 = E[\mathrm{tr}\nabla_X\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t] = E\bigg[\sum_{a=1}^m\sum_{b=1}^p \mathrm{d}_{ab}^X\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab}\bigg]$$

if $\lim_{X_{ab}\to\pm\infty}\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}h_{m,p}(\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{\Sigma}) = 0$ and

$$E[|(X_{ab}-\theta_{ab})\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}|] < \infty \tag{6.11}$$

for $a, c = 1, \dots, m$ and $b, d = 1, \dots, p$. Lemma 6.4 is used to obtain

$$E_1 = E\bigg[\mathrm{tr}\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\cdot\mathrm{tr}\boldsymbol{S}\boldsymbol{S}^+ + \sum_{a=1}^m\sum_{b=1}^p\{D_{ab}^{X\cdot 1}(\boldsymbol{\Phi}) + D_{ab}^{X\cdot 2}(\boldsymbol{\Phi})\}\bigg], \tag{6.12}$$

where $D_{ab}^{X\cdot 1}(\boldsymbol{\Phi})$ and $D_{ab}^{X\cdot 2}(\boldsymbol{\Phi})$ are defined in Lemma 6.4. Since

$$\sum_{a=1}^m\sum_{b=1}^p A_{1\cdot ab}^{jj}A_{2\cdot ab}^{ii} = 4\{\boldsymbol{F}\}_{ij}, \qquad \sum_{a=1}^m\sum_{b=1}^p A_{1\cdot ab}^{ij}A_{2\cdot ab}^{ij} = f_i + f_j + 2\{\boldsymbol{F}\}_{ij},$$

we observe that

$$\begin{aligned} \sum_{a=1}^m\sum_{b=1}^p D_{ab}^{X\cdot 1}(\boldsymbol{\Phi}) &= 2\sum_{i=1}^\ell f_i\frac{\partial\phi_i}{\partial f_i}, \\ \sum_{a=1}^m\sum_{b=1}^p D_{ab}^{X\cdot 2}(\boldsymbol{\Phi}) &= (m-\ell)\mathrm{tr}\boldsymbol{\Phi} + \sum_{i=1}^\ell\sum_{\substack{j=1\\ j\neq i}}^\ell\frac{(f_i+f_j)\phi_i}{f_i-f_j} \\ &= (m-\ell)\sum_{i=1}^\ell\phi_i + \sum_{i=1}^\ell\sum_{\substack{j=1\\ j\neq i}}^\ell\frac{2f_i\phi_i}{f_i-f_j} - \sum_{i=1}^\ell\sum_{\substack{j=1\\ j\neq i}}^\ell\phi_i \\ &= (m-2\ell+1)\sum_{i=1}^\ell\phi_i + 2\sum_{i=1}^\ell\sum_{j>i}^\ell\frac{f_i\phi_i - f_j\phi_j}{f_i-f_j}, \end{aligned}$$

which are substituted into (6.12) to obtain

$$E_1 = E\Big[\sum_{i=1}^{\ell}\Big\{b_{p,m,n}\phi_i + 2f_i\frac{\partial \phi_i}{\partial f_i} + 2\sum_{j>i}^{\ell}\frac{f_i\phi_i - f_j\phi_j}{f_i - f_j}\Big\}\Big], \tag{6.13}$$

where $b_{p,m,n} = n\wedge p + m - 2\ell + 1$. A simple manipulation gives that

$$\begin{aligned} b_{p,m,n} &= n\wedge p + m - (2m)\wedge\{2(n\wedge p)\} + 1 = n\wedge p + m + (-2m)\vee\{-2(n\wedge p)\} + 1 \\ &= (n\wedge p - m)\vee(m - n\wedge p) + 1 \\ &= |m - n\wedge p| + 1. \end{aligned}$$

Similarly, the Stein identity (6.1) can be used to rewrite E_2 as

$$E_2 = E[\mathrm{tr}\nabla_Y^t \boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+] = E\Big[\sum_{a=1}^{n}\sum_{b=1}^{p}\mathrm{d}_{ab}^Y\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{ab}\Big]$$

if $\lim_{Y_{ab}\to\pm\infty}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}h_{n,p}(\boldsymbol{Y}|\boldsymbol{0}_{n\times p},\boldsymbol{\Sigma}) = 0$ and

$$E[|Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}|] < \infty \tag{6.14}$$

for $a, c = 1, \ldots, n$ and $b, d = 1, \ldots, p$. Using Lemma 6.5, we get

$$E_2 = E\Big[\sum_{a=1}^{n}\sum_{b=1}^{p}\{D_{ab}^{Y\cdot 1}(\boldsymbol{\Phi}) + D_{ab}^{Y\cdot 2}(\boldsymbol{\Phi}) + D_{ab}^{Y\cdot 3}(\boldsymbol{\Phi})\}\Big]$$

It is here seen that

$$\sum_{a=1}^{n}\sum_{b=1}^{p}D_{ab}^{Y\cdot 1}(\boldsymbol{\Phi}) = \{n + p - 2(n\wedge p) - 1\}\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2.$$

For $B_{1\cdot ab}^{ij}$'s and $B_{2\cdot cd}^{ij}$'s given in Lemmas 6.3 and 6.5, it follows that

$$\sum_{a=1}^{n}\sum_{b=1}^{p}B_{1\cdot ab}^{jj}B_{2\cdot ab}^{ii} = -4\{\boldsymbol{F}\}_{ij}^2, \quad \sum_{a=1}^{n}\sum_{b=1}^{p}B_{1\cdot ab}^{ij}B_{2\cdot ab}^{ij} = -2f_if_j - 2\{\boldsymbol{F}\}_{ij}^2,$$

which yields that

$$\begin{aligned} \sum_{a=1}^{n}\sum_{b=1}^{p}D_{ab}^{Y\cdot 2}(\boldsymbol{\Phi}) &= -4\sum_{i=1}^{\ell}f_i^2\phi_i\frac{\partial\phi_i}{\partial f_i}, \\ \sum_{a=1}^{n}\sum_{b=1}^{p}D_{ab}^{Y\cdot 3}(\boldsymbol{\Phi}) &= -2\sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{f_if_j\phi_i^2}{f_i - f_j} = -2\sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}\frac{f_i^2\phi_i^2}{f_i - f_j} + 2\sum_{i=1}^{\ell}\sum_{\substack{j=1\\ j\neq i}}^{\ell}f_i\phi_i^2 \\ &= -2\sum_{i=1}^{\ell}\sum_{j>i}^{\ell}\frac{f_i^2\phi_i^2 - f_j^2\phi_j^2}{f_i - f_j} + 2(\ell - 1)\sum_{i=1}^{\ell}f_i\phi_i^2. \end{aligned}$$

Thus we obtain

$$E_2 = E\bigg[\sum_{i=1}^{\ell}\bigg\{a_{p,m,n}f_i\phi_i^2 - 4f_i^2\phi_i\frac{\partial\phi_i}{\partial f_i} - 2\sum_{j>i}^{\ell}\frac{f_i^2\phi_i^2 - f_j^2\phi_j^2}{f_i - f_j}\bigg\}\bigg], \tag{6.15}$$

where $a_{p,m,n} = n + p - 2(n \wedge p) + 2\ell - 3$. It is observed that

$$\begin{aligned} a_{p,m,n} &= n + p - 2(n \wedge p) + (2m) \wedge \{2(n \wedge p)\} - 3 \\ &= (n + p - 2(n \wedge p) + 2m) \wedge \{n + p - 2(n \wedge p) + 2(n \wedge p)\} - 3 \\ &= (|n - p| + 2m) \wedge (n + p) - 3. \end{aligned}$$

Combining (6.10), (6.13) and (6.15) provides the expression of risk given in Theorem 4.1. It is noted that the conditions (6.11) and (6.14) are satisfied when $E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2] < \infty$, which is proved in Lemma 6.6 given below. Thus the proof of Theorem 4.1 is complete. □

Lemma 6.6 *A sufficient condition for* (6.11) *and* (6.14) *is that* $E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2] < \infty$.

Proof. The Schwarz inequality leads to

$$\{E[|(X_{ab} - \theta_{ab})\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}|]\}^2 \le E[(X_{ab} - \theta_{ab})^2]E[\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}^2].$$

It is noted that $E[(X_{ab} - \theta_{ab})^2] < \infty$ and

$$\begin{aligned} E[\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}^2] &\le \sum_{c=1}^{m}\sum_{d=1}^{p} E[\{\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}^2] \\ &= E[\mathrm{tr}\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}^t] \\ &\le E[\mathrm{tr}\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t \cdot \mathrm{tr}\boldsymbol{S}] \\ &= E[\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2 \cdot \mathrm{tr}\boldsymbol{S}]. \end{aligned}$$

Hence (6.11) follows when $E[\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2 \cdot \mathrm{tr}\boldsymbol{S}] < \infty$.

Next, recalling that $\boldsymbol{X}\boldsymbol{H} = \boldsymbol{R}\boldsymbol{F}^{1/2}\boldsymbol{V}^t\boldsymbol{L}^{1/2}$, we see that

$$\begin{aligned} &E[|Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}|] \\ &= E\bigg[|Y_{ab}|\bigg|\sum_{i=1}^{\ell}\{\boldsymbol{Y}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}\}_{ci}f_i\phi_i^2\{\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\}_{id}\bigg|\bigg] \\ &\le E\bigg[|Y_{ab}|\sum_{i=1}^{\ell}|\{\boldsymbol{Y}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}\}_{ci}| \cdot f_i\phi_i^2 \cdot |\{\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\}_{id}|\bigg]. \end{aligned} \tag{6.16}$$

Since $\left(\sum_{i=1}^p a_i\right)^2 \le p\sum_{i=1}^p a_i^2$ for any set of real numbers $a_1, \ldots, a_p$, it is observed that

$$|Y_{ab}| \le \sum_{a=1}^{n}\sum_{b=1}^{p}|Y_{ab}| \le \sqrt{np\sum_{a=1}^{n}\sum_{b=1}^{p}Y_{ab}^2} = \sqrt{np\,\mathrm{tr}\boldsymbol{Y}^t\boldsymbol{Y}} = \sqrt{np\,\mathrm{tr}\boldsymbol{S}}. \tag{6.17}$$

Similarly, it is seen that

$$|\{\boldsymbol{Y}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}\}_{ci}| \le \sqrt{n\ell\mathrm{tr}\boldsymbol{V}^t\boldsymbol{L}^{-1/2}\boldsymbol{H}^t\boldsymbol{Y}^t\boldsymbol{Y}\boldsymbol{H}\boldsymbol{L}^{-1/2}\boldsymbol{V}} = \sqrt{n\ell^2}, \tag{6.18}$$

$$|\{\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\}_{id}| \le \sqrt{p\ell\mathrm{tr}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{V}\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t} \le \sqrt{p\ell\,\mathrm{tr}\boldsymbol{S}}. \tag{6.19}$$

Combining (6.16), (6.17), (6.18) and (6.19) yields that

$$E[|Y_{ab}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cd}|] \le np\ell^{3/2}E\left[(\mathrm{tr}\boldsymbol{S})\sum_{i=1}^{\ell} f_i\phi_i^2\right] = np\ell^{3/2}E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}\boldsymbol{\Phi}^2].$$

The sufficient condition for (6.14) is that the last r.h.s. given above is finite. Hence the proof is complete. □

6.2 Conditions for application of the Stein identity

The modified Stein type estimator is expressed as $\boldsymbol{\delta}^{mST} = \boldsymbol{X} - \boldsymbol{R}\boldsymbol{\Phi}^{mST}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$, where

$$\boldsymbol{\Phi}^{mST} = \mathrm{diag}(\phi_1^{mST}, \ldots, \phi_\ell^{mST}), \quad \phi_i^{mST} = \frac{c_i}{f_i} + \frac{d}{\mathrm{tr}\boldsymbol{F}},$$

where c_i's and d are positive constants and $c_1 \ge \cdots \ge c_\ell$. For the modified Stein type estimator $\boldsymbol{\delta}^{mST}$, the conditions (i), (ii) and (iii) of Theorem 4.1 are rewritten as follows:

(i) $E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}(\boldsymbol{\Phi}^{mST})^2] < \infty$,

(ii) $\lim_{X_{ab}\to\pm\infty}\{\boldsymbol{R}\boldsymbol{\Phi}^{mST}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cb}h_{m,p}(\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{\Sigma}) = 0$ for $a, c = 1, \ldots, m$ and $b, d = 1, \ldots, p$,

(iii) $\lim_{Y_{ab}\to\pm\infty}\{\boldsymbol{Y}\boldsymbol{S}^+\boldsymbol{X}^t\boldsymbol{R}(\boldsymbol{\Phi}^{mST})^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+\}_{cb}h_{n,p}(\boldsymbol{Y}|\boldsymbol{0}_{n\times p},\boldsymbol{\Sigma}) = 0$ for $a, c = 1, \ldots, n$ and $b, d = 1, \ldots, p$.

We can easily verify (ii) and (iii). A sufficient condition for (i) will here be established so that the Stein identity is applied to the risk of $\boldsymbol{\delta}^{mST}$. To this end, we provide useful lemmas.

Lemma 6.7 *If $\boldsymbol{X} \sim \mathcal{N}_{m\times p}(\boldsymbol{\Xi}, \boldsymbol{I}_m \otimes \boldsymbol{I}_p)$ with $m \ge p$, then we have*

$$E[\mathrm{tr}(\boldsymbol{X}^t\boldsymbol{X})^{-1}] \le p(m-p-1)^{-1} \quad \text{for } m-p \ge 2.$$

Proof. See the proof of Theorem 2 in Nishii and Krishnaiah (1988). □

Lemma 6.8 *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be defined as in the model* (1.1). *Denote $\boldsymbol{S} = \boldsymbol{Y}^t\boldsymbol{Y}$ and $\ell = m \wedge n \wedge p$. Then it follows that for $|m - n \wedge p| \ge 2$*

$$E[\mathrm{tr}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+|\boldsymbol{S}] \le \ell(|m - n\wedge p| - 1)^{-1}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}. \tag{6.20}$$

Proof. It is recalled that $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$ is the eigenvalue decomposition, where $\boldsymbol{L} = \mathrm{diag}(l_1, \ldots, l_{n\wedge p}) \in \mathbb{D}^+_{n\wedge p}$ and $\boldsymbol{H} \in \mathcal{V}_{p,n\wedge p}$. The joint p.d.f. of $(\boldsymbol{L}, \boldsymbol{H})$ is given by Muirhead (1982, Theorem 3.2.18) for the case of $n \geq p$ and Srivastava (2003, pp.1549) for the case of $p > n$. The joint p.d.f. can be expressed as

$$f(\boldsymbol{L}, \boldsymbol{H}) = K_0 \exp\Big(-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t\Big) \prod_{i=1}^{n\wedge p} l_i^{(|n-p|-1)/2} \prod_{i<j}^{n\wedge p}(l_i - l_j),$$

where K_0 is a normalizing constant. For the normalizing constant, see Muirhead (1982) and Srivastava (2003).

Note that

$$\begin{aligned}\mathrm{tr}(\boldsymbol{X} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\Theta})^t &= \mathrm{tr}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t \\ &\quad + 2\mathrm{tr}\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t + \mathrm{tr}\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}^t,\end{aligned}$$

where $[\boldsymbol{H}, \boldsymbol{H}_0] \in \mathcal{O}(p)$. Make the change of variables $(\boldsymbol{Z}, \boldsymbol{Z}_0) = (\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2}, \boldsymbol{X}\boldsymbol{H}_0)$. Since the Jacobian of the transformation is given by $J[\boldsymbol{X} \to (\boldsymbol{Z}, \boldsymbol{Z}_0)] = \prod_{i=1}^{n\wedge p} l_i^{m/2}$, the joint p.d.f. of $(\boldsymbol{Z}, \boldsymbol{Z}_0, \boldsymbol{L}, \boldsymbol{H})$ is proportional to

$$\begin{aligned}&\exp\Big(-\frac{1}{2}\mathrm{tr}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t - \mathrm{tr}\boldsymbol{Z}\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t \\ &\quad - \frac{1}{2}\mathrm{tr}\boldsymbol{Z}\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{Z}^t - \frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t\Big) \prod_{i=1}^{n\wedge p} l_i^{(|n-p|+m-1)/2} \prod_{i<j}^{n\wedge p}(l_i - l_j), \quad (6.21)\end{aligned}$$

where a normalizing constant is omitted, which implies that

$$\boldsymbol{Z}|\boldsymbol{Z}_0, \boldsymbol{S} \sim \mathcal{N}_{m\times(n\wedge p)}(\Xi, \boldsymbol{I}_m \otimes \boldsymbol{\Omega})$$

with $\Xi = -(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{\Omega}$ and $\boldsymbol{\Omega} = (\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}^{1/2})^{-1}$.

It is seen that $\boldsymbol{Z}$ is an $m \times (n \wedge p)$ full rank matrix and

$$\mathrm{tr}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+ = \mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}^t)^+ = \begin{cases}\mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}^t)^{-1} & \text{for } n\wedge p \geq m,\\ \mathrm{tr}(\boldsymbol{Z}^t\boldsymbol{Z})^{-1} & \text{for } n\wedge p < m.\end{cases}$$

Since $\boldsymbol{I}_{n\wedge p} \geq \boldsymbol{\Omega}^{-1}/(\mathrm{tr}\boldsymbol{\Omega}^{-1}) = \boldsymbol{\Omega}^{-1}/(\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S})$, we get

$$\mathrm{tr}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+ \leq \begin{cases}\mathrm{tr}(\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t)^{-1}(\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) & \text{for } n\wedge p \geq m,\\ \mathrm{tr}\boldsymbol{\Omega}(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}(\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) & \text{for } n\wedge p < m.\end{cases}$$

Using Lemma 6.7, we obtain (6.20) for $|m - n\wedge p| \geq 2$, which completes the proof. □

The l.h.s. of (i), namely $E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}(\boldsymbol{\Phi}^{mST})^2]$, is bounded above by

$$E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}(\boldsymbol{\Phi}^{mST})^2] \leq (c_1 + d)^2 E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{F}^{-1}] = (c_1 + d)^2 E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+]$$

since $c_1 \geq \cdots \geq c_\ell$ and $(\mathrm{tr}\boldsymbol{F})^{-1} \leq f_i^{-1}$ for every i. Using Lemma 6.8 gives that

$$E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}(\boldsymbol{X}\boldsymbol{S}^{+}\boldsymbol{X}^{t})^{+}] \leq \ell(|m - n \wedge p| - 1)^{-1} E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}]$$

for

$$|m - n \wedge p| \geq 2. \tag{6.22}$$

It is noted that $E[(\mathrm{tr}\boldsymbol{S})\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}]$ is always finite. Hence, under the condition (6.22), the Stein identity can be applied to the risk of the modified Stein type estimator $\boldsymbol{\delta}^{mST}$. When the condition (6.22) is met, we also make it possible to apply the Stein identity to the risks of the Efron-Morris type estimator $\boldsymbol{\delta}^{EM}$ and the modified Efron-Morris type estimator $\boldsymbol{\delta}^{mEM}$.

6.3 Proof of Theorem 5.1

For the proof, we use a similar argument as in Tsukuma (2010). Abbreviate $\boldsymbol{\Psi}(\boldsymbol{F})$ and $\boldsymbol{\Psi}_{+}(\boldsymbol{F})$ by $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \ldots, \psi_\ell)$ and $\boldsymbol{\Psi}_{+} = \mathrm{diag}(\psi_1^{+}, \ldots, \psi_\ell^{+})$, respectively. Let $\boldsymbol{H}_0$ be a $p \times (p - n \wedge p)$ matrix such that $[\boldsymbol{H}, \boldsymbol{H}_0] \in \mathcal{O}(p)$. It is observed that $\boldsymbol{\delta}^{SH} = \boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t + \boldsymbol{R}\boldsymbol{\Psi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t$ and

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\delta}^{SH} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\delta}^{SH} - \boldsymbol{\Theta})^t = {} & \mathrm{tr}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t \\ & + 2\mathrm{tr}\boldsymbol{\Psi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t\boldsymbol{R} \\ & + \mathrm{tr}\boldsymbol{\Psi}^2\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{R}.\end{aligned}$$

Thus the difference in risk of $\boldsymbol{\delta}_{+}^{SH}$ and $\boldsymbol{\delta}^{SH}$ is given by

$$\begin{aligned}& R(\boldsymbol{\delta}_{+}^{SH}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{SH}, \boldsymbol{\Theta}|\boldsymbol{\Sigma}) \\ & = E[\mathrm{tr}(\boldsymbol{\Psi}_{+}^2 - \boldsymbol{\Psi}^2)\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{R}] \\ & \quad + 2E[\mathrm{tr}(\boldsymbol{\Psi}_{+} - \boldsymbol{\Psi})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t\boldsymbol{R}].\end{aligned} \tag{6.23}$$

The first expectation in the r.h.s. of (6.23) is not positive because $(\psi_i^{+})^2 \leq \psi_i^2$ for all i.

Here, we use the same notation as in the proof of Lemma 6.8. Let $(\boldsymbol{Z}, \boldsymbol{Z}_0) = (\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2}, \boldsymbol{X}\boldsymbol{H}_0)$, where $[\boldsymbol{H}, \boldsymbol{H}_0] \in \mathcal{O}(p)$. Note that the joint p.d.f. of $(\boldsymbol{Z}, \boldsymbol{Z}_0, \boldsymbol{L}, \boldsymbol{H})$ is given by (6.21). Then the second expectation in the r.h.s. of (6.23) is expressed as

$$\iiint_{\mathbb{R}^{m\times(p-n\wedge p)}\times\mathbb{D}_{n\wedge p}^{+}\times\mathcal{V}_{p,n\wedge p}} I \times f(\boldsymbol{Z}_0, \boldsymbol{L}, \boldsymbol{H})(\mathrm{d}\boldsymbol{Z}_0)(\mathrm{d}\boldsymbol{L})(\mathrm{d}\boldsymbol{H}),$$

where

$$\begin{aligned}I = \int_{\mathbb{R}^{m\times(n\wedge p)}} & \mathrm{tr}(\boldsymbol{\Psi}_{+} - \boldsymbol{\Psi})\boldsymbol{R}^t\boldsymbol{Z}\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t\boldsymbol{R} \\ & \times \exp\Big(-\mathrm{tr}\boldsymbol{Z}\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t - \frac{1}{2}\mathrm{tr}\boldsymbol{Z}\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{Z}^t\Big)(\mathrm{d}\boldsymbol{Z})\end{aligned}$$

and

$$f(\boldsymbol{Z}_0, \boldsymbol{L}, \boldsymbol{H}) = K_1 \exp\Big(-\frac{1}{2}\mathrm{tr}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t - \frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t\Big) \times \prod_{i=1}^{n\wedge p} l_i^{(|n-p|+m-1)/2} \prod_{i<j}^{n\wedge p}(l_i - l_j)$$

with a normalizing constant K_1. Hence, if it is shown that $I \le 0$, the proof of Theorem 5.1 will be complete.

We next consider the singular value decomposition $\boldsymbol{Z} = \boldsymbol{R}\boldsymbol{D}\boldsymbol{V}^t$, where $\boldsymbol{R} \in \mathcal{V}_{m,\ell}$, $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_\ell) = \boldsymbol{F}^{1/2} \in \mathbb{D}_\ell^+$, $\boldsymbol{V} \in \mathcal{V}_{n\wedge p,\ell}$ and $\ell = m \wedge (n \wedge p)$. From Theorem 5 of Uhlig (1994), the Jacobian of the transformation $\boldsymbol{Z} = \boldsymbol{R}\boldsymbol{D}\boldsymbol{V}^t$ is given by

$$\begin{aligned}(\mathrm{d}\boldsymbol{Z}) &= \frac{1}{2^\ell}\prod_{i=1}^{\ell} d_i^{|m-n\wedge p|}\prod_{i<j}^{\ell}(d_i^2 - d_j^2)(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})(\mathrm{d}\boldsymbol{D})(\boldsymbol{V}^t\mathrm{d}\boldsymbol{V}) \\ &= \frac{1}{2^{2\ell}}\prod_{i=1}^{\ell} f_i^{(|m-n\wedge p|-1)/2}\prod_{i<j}^{\ell}(f_i - f_j)(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})(\mathrm{d}\boldsymbol{F})(\boldsymbol{V}^t\mathrm{d}\boldsymbol{V}),\end{aligned}$$

where the second equality is verified by the transformation $\boldsymbol{F} = \boldsymbol{D}^2$. Note that $(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})$ and $(\boldsymbol{V}^t\mathrm{d}\boldsymbol{V})$ are invariant with respect to an orthogonal transformation (Muirhead (1982, pp.69)). For $i = 1, \ldots, \ell$, it is observed that

$$\{\boldsymbol{R}^t\boldsymbol{Z}\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t\boldsymbol{R}\}_{ii} = f_i^{1/2}\boldsymbol{v}_i^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t - \boldsymbol{\Theta})^t\boldsymbol{r}_i = \boldsymbol{a}_i^t\boldsymbol{r}_i, \quad \text{say},$$

where $\boldsymbol{v}_i$ and $\boldsymbol{r}_i$ are the i-th column vectors of $\boldsymbol{V}$ and $\boldsymbol{R}$, respectively. We then obtain

$$I = \sum_{i=1}^{\ell}\iiint_{\mathcal{V}_{m,\ell}\times\mathbb{D}_\ell^+\times\mathcal{V}_{n\wedge p,\ell}}(\psi_i^+ - \psi_i)\boldsymbol{a}_i^t\boldsymbol{r}_i e^{-\boldsymbol{a}_i^t\boldsymbol{r}_i}G_i(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})(\mathrm{d}\boldsymbol{F})(\boldsymbol{V}^t\mathrm{d}\boldsymbol{V}), \tag{6.24}$$

where

$$G_i = \exp\Big\{-\sum_{j\neq i}\boldsymbol{a}_j^t\boldsymbol{r}_j - \frac{1}{2}\mathrm{tr}\boldsymbol{F}\boldsymbol{V}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{V}\Big\}\frac{1}{2^{2\ell}}\prod_{i=1}^{\ell} f_i^{(|m-n\wedge p|-1)/2}\prod_{i<j}^{\ell}(f_i - f_j).$$

For each i, we make the transformation $\boldsymbol{r}_i \to -\boldsymbol{r}_i$. This transformation is equivalent to the orthogonal transformation $\boldsymbol{R} \to \boldsymbol{R}\boldsymbol{O}_i$, where $\boldsymbol{O}_i \in \mathcal{O}(\ell)$ such that the i-th diagonal is minus one and the other diagonals are ones. Because $(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})$ is invariant with respect to the orthogonal transformation, (6.24) is rewritten as

$$I = \sum_{i=1}^{\ell}\iiint_{\mathcal{V}_{m,\ell}\times\mathbb{D}_\ell^+\times\mathcal{V}_{n\wedge p,\ell}}(\psi_i^+ - \psi_i)(-\boldsymbol{a}_i^t\boldsymbol{r}_i e^{\boldsymbol{a}_i^t\boldsymbol{r}_i})G_i(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})(\mathrm{d}\boldsymbol{F})(\boldsymbol{V}^t\mathrm{d}\boldsymbol{V}). \tag{6.25}$$

Adding each side of (6.24) and (6.25) yields that

$$2I = \sum_{i=1}^{\ell}\iiint_{\mathcal{V}_{m,\ell}\times\mathbb{D}_\ell^+\times\mathcal{V}_{n\wedge p,\ell}}(\psi_i^+ - \psi_i)\boldsymbol{a}_i^t\boldsymbol{r}_i(e^{-\boldsymbol{a}_i^t\boldsymbol{r}_i} - e^{\boldsymbol{a}_i^t\boldsymbol{r}_i})G_i(\boldsymbol{R}^t\mathrm{d}\boldsymbol{R})(\mathrm{d}\boldsymbol{F})(\boldsymbol{V}^t\mathrm{d}\boldsymbol{V}).$$

Since $\psi_i^+ \ge \psi_i$ and $\boldsymbol{a}_i^t\boldsymbol{r}_i(e^{-\boldsymbol{a}_i^t\boldsymbol{r}_i} - e^{\boldsymbol{a}_i^t\boldsymbol{r}_i}) \le 0$ for any value of $\boldsymbol{a}_i^t\boldsymbol{r}_i$, it always holds that $I \le 0$. Thus the proof of Theorem 5.1 is complete. □

Acknowledgments

The authors are deeply grateful to two anonymous reviewers and an associate editor for their helpful suggestions. The first author acknowledges support from Grant-in-Aid for Scientific Research (23500361), Japan. The second author acknowledges support from Grant-in-Aid for Scientific Research (21540114, 23243039 and 26330036), Japan.

References

[1] Baranchik, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **41**, 642–645.

[2] Chételat, D. and Wells, M.T. (2012). Improved multivariate normal mean estimation with unknown covariance when p is greater than n, *Ann. Statist.*, **40**, 3137–3160.

[3] Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika*, **59**, 335–347.

[4] Konno, Y. (1990). Families of minimax estimators of matrix of normal means with unknown covariance matrix. *J. Japan Statist. Soc.*, **20**, 191–201.

[5] Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.*, **36**, 44–55.

[6] Konno, Y. (1992). Improved estimation of matrix of normal mean and eigenvalues in the multivariate F-distribution, Doctoral dissertation, Institute of Mathematics, University of Tsukuba. This is downloadable from his website (http://mcm-www.jwu.ac.jp/˜konno/).

[7] Konno, Y. (2009). Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss, *J. Multivariate Anal.*, **100**, 2237–2253.

[8] Kubokawa, T. and Srivastava, M.S. (2001). Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. *J. Multivariate Anal.*, **76**, 138–152.

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

[10] Nishii, R. and Krishnaiah, P, R. (1988). On the moments of classical estimates of explanatory variables under a multivariate calibration model, *Sankhyā*, **50**, 137–148.

[11] Srivastava, M.S. (2003). Singular Wishart and multivariate beta distributions, *Ann. Statist.*, **31**, 1537–1560.

[12] Stein, C. (1973). Estimation of the mean of a multivariate normal distribution, In *Proc. Prague Symp. Asymptotic Statist.*, 345–381.

[13] Tsukuma, H. (2010). Shrinkage minimax estimation and positive-part rule for a mean matrix in an elliptically contoured distribution, *Statist. & Probab. Letters*, **80**, 215–220.

[14] Tsukuma, H. and Kubokawa, T. (2007). Methods for improvement in estimation of a normal mean matrix, *J. Multivariate Anal.*, **98**, 1592–1610.

[15] Uhlig, H. (1994). On singular Wishart and singular multivariate Beta distributions, *Ann. Statist.*, **22**, 395–405.