CIRJE-F-930

Estimation of the Mean Vector in a Singular Multivariate Normal Distribution

Hisayuki Tsukuma
Toho University

Tatsuya Kubokawa
The University of Tokyo

April 2014

Estimation of the Mean Vector in a Singular Multivariate Normal Distribution

Hisayuki Tsukuma* and Tatsuya Kubokawa†

April 23, 2014

Abstract

This paper addresses the problem of estimating the mean vector of a singular multivariate normal distribution with an unknown singular covariance matrix. The maximum likelihood estimator is shown to be minimax relative to a quadratic loss weighted by the Moore-Penrose inverse of the covariance matrix. An unbiased risk estimator relative to the weighted quadratic loss is provided for a Baranchik type class of shrinkage estimators. Based on the unbiased risk estimator, a sufficient condition for the minimaxity is expressed not only as a differential inequality, but also as an integral inequality. Also, generalized Bayes minimax estimators are established by using an interesting structure of singular multivariate normal distribution.

AMS 2010 subject classifications: Primary 62C20, 62J07; secondary 62C10.

Key words and phrases: Empirical Bayes method, generalized Bayes estimator, inadmissibility, minimaxity, Moore-Penrose inverse, pseudo-Wishart distribution, quadratic loss, shrinkage estimator, statistical decision theory.

1 Introduction

Statistical inference with the determinant and the inverse of sample covariance matrix requires nonsingularity of the sample covariance matrix. However in practical cases of data analysis, the nonsingularity is not always satisfied. The singularity occurs for many reasons, but in general such singularity is very hard to handle. This paper treats a singular multivariate normal model, which yields a singular sample covariance matrix, and aims to provide a series of decision-theoretic results in estimation of the mean vector.

*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

The singular multivariate normal distribution model and the related topics have been studied for a long time in the literature. For the density function, see Khatri (1968), Rao (1973) and Srivastava and Khatri (1979). Khatri (1968) and Rao (1973) derived the maximum likelihood estimators for the mean vector and the singular covariance matrix. Srivastava (2003) and Díaz-García, *et al.* (1997) studied central and noncentral pseudo-Wishart distributions which have been used for developing distribution theories in the problems of testing hypotheses. However, little is known about a decision-theoretic approach to estimation in the singular model.

To specify the singular model addressed in this paper, let $\boldsymbol{X}$ and $\boldsymbol{Y}_i$ $(i = 1, \dots, n)$ be p-dimensional random vectors having the stochastic representations

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{Z}_0, \\ \boldsymbol{Y}_i &= \boldsymbol{B}\boldsymbol{Z}_i, \quad i = 1, \dots, n, \end{aligned} \tag{1.1}$$

where $\boldsymbol{Z}_0, \boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ are mutually and independently distributed as $\mathcal{N}_r(\boldsymbol{0}_r, \boldsymbol{I}_r)$, and $\boldsymbol{\theta}$ and $\boldsymbol{B}$ are, respectively, a p-dimensional vector and a $p \times r$ matrix of unknown parameters. Then we write $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and $\boldsymbol{Y}_i \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ $(i = 1, \dots, n)$, where $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^t$. Assume that

$$r \le \min(n, p),$$

and $\boldsymbol{B}$ is of full column rank, namely $\boldsymbol{\Sigma}$ is a positive semi-definite matrix of rank r. In the case when $r < p$, technically speaking, $\mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is called the singular multivariate normal distribution with mean vector $\boldsymbol{\theta}$ and singular covariance $\boldsymbol{\Sigma}$. For the definition of the singular multivariate normal distribution, see Khatri (1968), Rao (1973, Chapter 8) and Srivastava and Khatri (1979, page 43).

Denote by $\boldsymbol{\Sigma}^+$ the Moore-Penrose inverse of $\boldsymbol{\Sigma}$. Consider the problem of estimating the mean vector $\boldsymbol{\theta}$ relative to quadratic loss weighted by $\boldsymbol{\Sigma}^+$,

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = (\boldsymbol{\delta} - \boldsymbol{\theta})^t \boldsymbol{\Sigma}^+ (\boldsymbol{\delta} - \boldsymbol{\theta}), \tag{1.2}$$

where $\boldsymbol{\delta}$ is an estimator of $\boldsymbol{\theta}$ based on $\boldsymbol{X}$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n)^t$. The accuracy of estimators is compared by the risk function $R(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E[L(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma})]$, where the expectation is taken with respect to (1.1).

A natural estimator of $\boldsymbol{\theta}$ is the unbiased estimator $\boldsymbol{\delta}^{UB} = \boldsymbol{X}$, which is also the maximum likelihood estimator as pointed out by Khatri (1968, page 276) and Rao (1973, page 532). This paper considers improvement on $\boldsymbol{\delta}^{UB}$ via the Baranchik (1970) type class of shrinkage estimators

$$\boldsymbol{\delta}^{SH} = \Big(1 - \frac{\phi(F)}{F}\Big)\boldsymbol{X}, \qquad F = \boldsymbol{X}^t \boldsymbol{S}^+ \boldsymbol{X},$$

where $\phi(F)$ is a bounded and differentiable function of F.

It is worth noting that, instead of F in $\boldsymbol{\delta}^{SH}$, we may use $F^- = \boldsymbol{X}^t \boldsymbol{S}^- \boldsymbol{X}$, where $\boldsymbol{S}^-$ is a generalized inverse of $\boldsymbol{S}$. Since the generalized inverse is not unique, it may be troublesome to consider which we employ as the generalized inverse. On the other hand, the Moore-Penrose inverse is unique and it is easy to discuss its distributional property. See Srivastava (2007) for interesting discussion on the Hotelling type T-square tests with the Moore Penrose and the generalized inverses in high dimension.

The rest of this paper is organized as follows. In Section 2, we introduce the definition of the Moore-Penrose inverse and its useful properties. We then set up a decision-theoretic framework for estimating $\boldsymbol{\theta}$ and derive some properties of estimators and their risk functions which are specific to the singular model. The key tool for their derivations is the equality

$$\boldsymbol{S}\boldsymbol{S}^{+} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{+},$$

which holds with probability one, where $\boldsymbol{S} = \boldsymbol{Y}^t\boldsymbol{Y}$ and $\boldsymbol{S}^{+}$ is the Moore-Penrose inverse of $\boldsymbol{S}$. In Section 2, we also prove the minimaxity of $\boldsymbol{\delta}^{UB}$. In Section 3, we obtain sufficient conditions for the minimaxity of $\boldsymbol{\delta}^{SH}$. These conditions are given not only by a differential inequality, but also by an integral inequality. In Section 4, an empirical Bayes motivation is given for the James-Stein (1961) type shrinkage estimator and its positive part estimator. Also, Section 4 suggests a hierarchical prior in the singular model and shows that the resulting generalized Bayes estimators are minimax. Section 5 provides some remarks on related topics.

2 Estimation in the Singular Normal Model

2.1 The Moore-Penrose inverse and its useful properties

We begin by introducing the following notations which will be used through the paper. Let $\mathcal{O}(r)$ be the group of orthogonal matrices of order r. For $p \geq r$, the Stiefel manifold is denoted by $\mathcal{V}_{p,r} = \{\boldsymbol{A} \in \mathbb{R}^{p\times r} : \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I}_r\}$. It is noted that $\mathcal{V}_{r,r} = \mathcal{O}(r)$. Let $\mathcal{D}_r$ be a set of $r \times r$ diagonal matrices whose diagonal elements $d_1, \ldots, d_r$ satisfy $d_1 > \cdots > d_r > 0$.

As an inverse matrix of a singular covariance matrix, we use the Moore-Penrose inverse matrix, which is defined as follows:

Definition 2.1 *For a matrix $\boldsymbol{A}$, there exists a matrix $\boldsymbol{A}^{+}$ such that* (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) $(\boldsymbol{A}^{+}\boldsymbol{A})^t = \boldsymbol{A}^{+}\boldsymbol{A}$. *Then $\boldsymbol{A}^{+}$ is called the Moore-Penrose inverse of $\boldsymbol{A}$.*

The following basic properties and results on the Moore-Penrose inverse matrix are useful for investigating properties of shrinkage estimators. Lemmas 2.1 and 2.2 are due to Harville (1997, Chapter 20).

Lemma 2.1 *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* $\boldsymbol{A}$.

Lemma 2.2 *Let* $\boldsymbol{B}$ *be a* $p \times r$ *matrix of full column rank. We then have*
(1) $\boldsymbol{B}^+\boldsymbol{B} = \boldsymbol{I}_r$,
(2) $\boldsymbol{B}\boldsymbol{B}^+$ *is idempotent,*
(3) $\boldsymbol{B}^+ = (\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t$, *in particular* $\boldsymbol{H}^+ = \boldsymbol{H}^t$ *for* $\boldsymbol{H} \in \mathcal{V}_{p,r}$,
(4) $(\boldsymbol{B}\boldsymbol{C}^t)^+ = (\boldsymbol{C}^t)^+\boldsymbol{B}^+ = \boldsymbol{C}(\boldsymbol{C}^t\boldsymbol{C})^{-1}(\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t$ *for a* $q \times r$ *matrix* $\boldsymbol{C}$ *of full column rank.*

Lemma 2.3 *Let* $\boldsymbol{A}$ *be an* $r \times r$ *nonsingular matrix and* $\boldsymbol{B}$ *a* $p \times r$ *matrix of full column rank. Then we have* $(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^t)^+ = (\boldsymbol{B}^t)^+\boldsymbol{A}^{-1}\boldsymbol{B}^+$. *In particular, it follows that* $(\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t)^+ = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t$ *for* $\boldsymbol{H} \in \mathcal{V}_{p,r}$ *and* $\boldsymbol{L} \in \mathcal{D}_r$.

Proof. It is noted that $\boldsymbol{A}\boldsymbol{B}^t$ is of full column rank. From (3) of Lemma 2.1 and (4) of Lemma 2.2, it follows that $(\boldsymbol{A}\boldsymbol{B}^t)^+ = (\boldsymbol{B}^t)^+\boldsymbol{A}^+ = (\boldsymbol{B}^t)^+\boldsymbol{A}^{-1}$ and

$$(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^t)^+ = \{\boldsymbol{B}(\boldsymbol{A}\boldsymbol{B}^t)\}^+ = (\boldsymbol{A}\boldsymbol{B}^t)^+\boldsymbol{B}^+ = (\boldsymbol{B}^t)^+\boldsymbol{A}^{-1}\boldsymbol{B}^+.$$

The second part immediately follows from (3) of Lemma 2.2. □

The Moore-Penrose inverse enables us to provide a general form of a solution of a homogeneous linear system. The following lemma is given in Harville (1997, Theorem 11.2.1).

Lemma 2.4 *Let* $\boldsymbol{x}$ *be a* p*-dimensional vector of unknown variables and* $\boldsymbol{A}$ *an* $r \times p$ *coefficient matrix. Then a solution of a homogeneous linear system* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}_r$ *is given by* $\boldsymbol{x}_0 = (\boldsymbol{I}_p - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{b}$ *for some* p*-dimensional vector* $\boldsymbol{b}$.

The following lemma is due to Zhang (1985). See also Olkin (1998).

Lemma 2.5 *Let* $\boldsymbol{B}$ *be a* $p \times r$ *random matrix of full column rank. Denote the density of* $\boldsymbol{B}$ *by* $f(\boldsymbol{B})$ *and the Moore-Penrose inverse of* $\boldsymbol{B}$ *by* $\boldsymbol{C} = \boldsymbol{B}^+ = (\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t$. *Then the density of* $\boldsymbol{C}$ *is given by* $|\boldsymbol{C}\boldsymbol{C}^t|^{-p}f(\boldsymbol{C}^+)$.

Next, we provide some remarks on the probability density function of singular multivariate normal distribution (1.1). An explicit expression of the density function is given as follows: Define $\boldsymbol{B}_0 \in \mathcal{V}_{p,p-r}$ such that $\boldsymbol{B}_0^+\boldsymbol{B} = \boldsymbol{B}_0^t\boldsymbol{B} = \boldsymbol{0}_{(p-r)\times r}$, namely $\boldsymbol{\Sigma}\boldsymbol{B}_0 = \boldsymbol{0}_{p\times(p-r)}$. Denote $\boldsymbol{\Lambda} \in \mathcal{D}_r$, where the diagonal elements of $\boldsymbol{\Lambda}$ consist of nonzero ordered eigenvalues of $\boldsymbol{\Sigma}$. Then the density of $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ with a singular $\boldsymbol{\Sigma}$ of rank r is expressed as

$$f(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = (2\pi)^{-r/2}|\boldsymbol{\Lambda}|^{-1/2}\exp\Big\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\theta})^t\boldsymbol{\Sigma}^+(\boldsymbol{X}-\boldsymbol{\theta})\Big\}, \tag{2.1}$$

where

$$\boldsymbol{B}_0^+(\boldsymbol{X}-\boldsymbol{\theta}) = \boldsymbol{0}_{p-r} \quad \text{with probability one.}$$

This density is interpreted as the density on the hyperplane $\boldsymbol{B}_0^+(\boldsymbol{X}-\boldsymbol{\theta})=\boldsymbol{0}_{p-r}$. The above expression is given by Khatri (1968), Rao (1973) and Srivastava and Khatri (1979), where their expression uses a generalized inverse of $\boldsymbol{\Sigma}$ instead of the Moore-Penrose inverse $\boldsymbol{\Sigma}^+$.

From Lemma 2.2 (4), it follows that $\boldsymbol{\Sigma}^+=(\boldsymbol{B}^t)^+\boldsymbol{B}^+$. It is also noted that $|\boldsymbol{I}_p-\lambda\boldsymbol{\Sigma}|=|\boldsymbol{I}_p-\lambda\boldsymbol{B}\boldsymbol{B}^t|=|\boldsymbol{I}_r-\lambda\boldsymbol{B}^t\boldsymbol{B}|$ for a scalar λ, so that the nonzero eigenvalues of $\boldsymbol{\Sigma}=\boldsymbol{B}\boldsymbol{B}^t$ are equivalent to those of a positive definite matrix $\boldsymbol{B}^t\boldsymbol{B}$. Hence (2.1) is alternatively expressed as

$$f(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{B})=(2\pi)^{-r/2}|\boldsymbol{B}^t\boldsymbol{B}|^{-1/2}\exp\Big\{-\frac{1}{2}\|\boldsymbol{B}^+(\boldsymbol{X}-\boldsymbol{\theta})\|^2\Big\}.$$

Similarly, the joint density of $\boldsymbol{Y}_i=\boldsymbol{B}\boldsymbol{Z}_i$ $(i=1,\dots,n)$ with $\boldsymbol{Z}_i\sim\mathcal{N}_r(\boldsymbol{0}_r,\boldsymbol{I}_r)$ is given by

$$\begin{aligned}f(\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_i|\boldsymbol{B})&=(2\pi)^{-nr/2}|\boldsymbol{B}^t\boldsymbol{B}|^{-n/2}\exp\Big\{-\frac{1}{2}\sum_{i=1}^n\|\boldsymbol{B}^+\boldsymbol{Y}_i\|^2\Big\}\\&=(2\pi)^{-nr/2}|\boldsymbol{B}^t\boldsymbol{B}|^{-n/2}\exp\Big\{-\frac{1}{2}\mathrm{tr}\,(\boldsymbol{B}^t)^+\boldsymbol{B}^+\boldsymbol{S}\Big\},\end{aligned}$$

where $\boldsymbol{S}=\boldsymbol{Y}^t\boldsymbol{Y}$ with $\boldsymbol{Y}=(\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n)^t$ and

$$\boldsymbol{B}_0^+\boldsymbol{Y}_i=\boldsymbol{0}_{p-r}\quad(i=1,\dots,n)\quad\text{with probability one.}$$

2.2 Risk properties under a quadratic loss with a singular weighted matrix

In this paper, we consider the estimation of the unknown mean vector $\boldsymbol{\theta}$ in the canonical model given by $\boldsymbol{X}\sim\mathcal{N}_p(\boldsymbol{\theta},\boldsymbol{\Sigma})$ and $\boldsymbol{Y}_i\sim\mathcal{N}_p(\boldsymbol{0}_p,\boldsymbol{\Sigma})$ $(i=1,\dots,n)$, where $\boldsymbol{X}$ and $\boldsymbol{Y}_i$'s are mutually independent and $\boldsymbol{\Sigma}$ is an unknown positive semi-definite matrix of rank r. It is assumed that

$$r\le\min(n,p).$$

An estimator $\boldsymbol{\delta}$ of $\boldsymbol{\theta}$ is evaluated by the risk function relative to the quadratic loss function with a singular weighted matrix given in (1.2), namely, $L(\boldsymbol{\delta},\boldsymbol{\theta}|\Sigma)=(\boldsymbol{\delta}-\boldsymbol{\theta})^t\boldsymbol{\Sigma}^+(\boldsymbol{\delta}-\boldsymbol{\theta})$.

Let $\boldsymbol{S}=\boldsymbol{Y}^t\boldsymbol{Y}$. Denote by $\boldsymbol{S}^+$ the Moore-Penrose inverse of $\boldsymbol{S}$. From the definition of the singular multivariate normal distribution, $\boldsymbol{Y}$ is expressed as

$$\boldsymbol{Y}=(\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n)^t=(\boldsymbol{B}\boldsymbol{Z}_1,\dots,\boldsymbol{B}\boldsymbol{Z}_n)^t=\boldsymbol{Z}\boldsymbol{B}^t$$

where $\boldsymbol{Z}=(\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n)^t$ with $\boldsymbol{Z}_i\sim\mathcal{N}_r(\boldsymbol{0}_r,\boldsymbol{I}_r)$. Since $r\le\min(n,p)$, the $n\times p$ matrix $\boldsymbol{Y}$ has rank r, and so does $\boldsymbol{S}$. Then, the following equality is useful for investigating a risk performance under the singular weighted quadratic loss:

$$\boldsymbol{S}\boldsymbol{S}^+=\boldsymbol{S}^+\boldsymbol{S}=\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+=\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}.\tag{2.2}$$

In fact, it is noted that $\boldsymbol{S}$ has the stochastic representation $\boldsymbol{S} = \boldsymbol{B}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{B}^t$ and hence $\boldsymbol{S}^+ = (\boldsymbol{B}^t)^+(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}\boldsymbol{B}^+$ from Lemma 2.3. It turns out that

$$\boldsymbol{S}\boldsymbol{S}^+ = \boldsymbol{B}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{B}^t(\boldsymbol{B}^t)^+(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}\boldsymbol{B}^+ = \boldsymbol{B}\boldsymbol{B}^+ = \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t = \Sigma\Sigma^+$$

with probability one, since $\boldsymbol{B}^+\boldsymbol{B} = \boldsymbol{B}^t(\boldsymbol{B}^t)^+ = \boldsymbol{I}_r$ from Lemma 2.1 (2) and Lemma 2.2 (1). Since $\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t$ is symmetric, it follows that $\boldsymbol{\Sigma}^+\boldsymbol{\Sigma} = \boldsymbol{S}^+\boldsymbol{S}$.

Proposition 2.1 *For any estimator $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{X}, \boldsymbol{S})$, the estimator $\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta}$ has the same risk as $\boldsymbol{\delta}$ under the singular weighted quadratic loss* (1.2) *if $r \le \min(n, p)$.*

Proof. The risk of $\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta}$ is given by

$$\begin{aligned} R(\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta}, \boldsymbol{\theta}|\Sigma) &= E[(\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta} - \boldsymbol{\theta})^t\Sigma^+(\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta} - \boldsymbol{\theta})] \\ &= E[\boldsymbol{\delta}^t\boldsymbol{S}^+\boldsymbol{S}\Sigma^+\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta} - 2\boldsymbol{\theta}^t\Sigma^+\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta} + \boldsymbol{\theta}^t\Sigma^+\boldsymbol{\theta}]. \end{aligned}$$

It follows from (2.2) that $\boldsymbol{S}^+\boldsymbol{S}\Sigma^+\boldsymbol{S}\boldsymbol{S}^+ = \Sigma^+\Sigma\Sigma^+\Sigma\Sigma^+ = \Sigma^+\Sigma\Sigma^+ = \Sigma^+$ and that $\Sigma^+\boldsymbol{S}\boldsymbol{S}^+ = \Sigma^+\Sigma\Sigma^+ = \Sigma^+$. Thus, $R(\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\delta}, \boldsymbol{\theta}|\Sigma) = E[\boldsymbol{\delta}^t\Sigma^+\boldsymbol{\delta} - 2\boldsymbol{\theta}^t\Sigma^+\boldsymbol{\delta} + \boldsymbol{\theta}^t\Sigma^+\boldsymbol{\theta}] = R(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma})$, which shows Proposition 2.1. $\square$

In estimation of a normal mean vector with $p = r > n$, Chételat and Wells (2012) considered a class of estimators,

$$\begin{aligned} \boldsymbol{\delta}^{CW} &= (\boldsymbol{I}_p - \psi(\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X})\boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X} \\ &= \{1 - \psi(\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X})\}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} + (\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}, \end{aligned}$$

where ψ is a scalar-valued function of $\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}$. When $r \le \min(n, p)$, it follows from Proposition 2.1 that $\boldsymbol{\delta}^{CW}$, $\{1 - \psi(\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X})\}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}$ and $\{1 - \psi(\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X})\}\boldsymbol{X}$ have the same risk function. In general, the estimator $g_1(\boldsymbol{X}, \boldsymbol{S})\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} + g_2(\boldsymbol{X}, \boldsymbol{S})(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}$ for nonnegative scalar-valued functions g_1 and g_2 has the same risk as the estimator $g_1(\boldsymbol{X}, \boldsymbol{S})\boldsymbol{X}$ in the case of $r \le \min(n, p)$.

It is well known that the James-Stein (1961) type estimator can be improved on by the positive-part James-Stein estimator (Baranchik (1970)). This dominance property can be extended to our situation. Consider a shrinkage estimator of the form $g(\boldsymbol{X}, \boldsymbol{S})\boldsymbol{X}$ for an integrable and scalar-valued function $g(\boldsymbol{X}, \boldsymbol{S})$.

Proposition 2.2 *Assume that the risk of $g(\boldsymbol{X}, \boldsymbol{S})\boldsymbol{X}$ is finite and that*

$$g(\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} + (\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}, \boldsymbol{S}) = g(-\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} + (\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}, \boldsymbol{S}). \tag{2.3}$$

If $\Pr(g(\boldsymbol{X}, \boldsymbol{S}) > 0) < 1$*, then $\boldsymbol{\delta} = g(\boldsymbol{X}, \boldsymbol{S})\boldsymbol{X}$ is dominated by*

$$\boldsymbol{\delta}^{TR} = g_+(\boldsymbol{X}, \boldsymbol{S})\boldsymbol{X}, \quad g_+(\boldsymbol{X}, \boldsymbol{S}) = \max\{0, g(\boldsymbol{X}, \boldsymbol{S})\},$$

relative to the loss (1.2).

For instance, the condition (2.3) is satisfied by $g(\boldsymbol{X},\boldsymbol{S})=\psi(\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X})$.

Proof of Proposition 2.2. Take $\boldsymbol{B}_0\in\mathcal{V}_{p,p-r}$ such that $\boldsymbol{B}_0^+\boldsymbol{B}=\boldsymbol{B}_0^t\boldsymbol{B}=\boldsymbol{0}_{(p-r)\times r}$. It follows by (2.1) that $\boldsymbol{B}_0^t\boldsymbol{X}=\boldsymbol{B}_0^t\boldsymbol{\theta}$ with probability one. Let $\boldsymbol{H}=\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}$, where $(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}=\{(\boldsymbol{B}^t\boldsymbol{B})^{1/2}\}^{-1}$ and $(\boldsymbol{B}^t\boldsymbol{B})^{1/2}$ is a symmetric square root of $\boldsymbol{B}^t\boldsymbol{B}$, namely $\boldsymbol{B}^t\boldsymbol{B}=(\boldsymbol{B}^t\boldsymbol{B})^{1/2}(\boldsymbol{B}^t\boldsymbol{B})^{1/2}$. Note that $\boldsymbol{H}\in\mathcal{V}_{p,r}$. Since $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+=\boldsymbol{H}\boldsymbol{H}^t$ and $\boldsymbol{I}_p-\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+=\boldsymbol{B}_0\boldsymbol{B}_0^t$, using the identity (2.2) yields that

$$\begin{aligned}\boldsymbol{X}&=\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}+(\boldsymbol{I}_p-\boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}\\&=\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\boldsymbol{X}+(\boldsymbol{I}_p-\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+)\boldsymbol{X}\\&=\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}+\boldsymbol{B}_0\boldsymbol{B}_0^t\boldsymbol{\theta}.\end{aligned}$$

Here, we abbreviate $g(\boldsymbol{X},\boldsymbol{S})$ and $g_+(\boldsymbol{X},\boldsymbol{S})$ by $g(\boldsymbol{H}^t\boldsymbol{X})$ and $g_+(\boldsymbol{H}^t\boldsymbol{X})$, respectively. The difference in risk between $\boldsymbol{\delta}^{TR}$ and $\boldsymbol{\delta}$ can be written as $R(\boldsymbol{\delta}^{TR},\boldsymbol{\theta}|\boldsymbol{\Sigma})-R(\boldsymbol{\delta},\boldsymbol{\theta}|\boldsymbol{\Sigma})=E_2-2E_1$, where

$$\begin{aligned}E_1&=E[\{g_+(\boldsymbol{H}^t\boldsymbol{X})-g(\boldsymbol{H}^t\boldsymbol{X})\}\boldsymbol{X}^t\boldsymbol{\Sigma}^+\boldsymbol{\theta}],\\E_2&=E[\{g_+^2(\boldsymbol{H}^t\boldsymbol{X})-g^2(\boldsymbol{H}^t\boldsymbol{X})\}\boldsymbol{X}^t\boldsymbol{\Sigma}^+\boldsymbol{X}].\end{aligned}$$

Since $g_+^2(\boldsymbol{H}^t\boldsymbol{X})\le g^2(\boldsymbol{H}^t\boldsymbol{X})$ for any $\boldsymbol{X}$ and $\boldsymbol{S}$, it follows that $E_2\le 0$. Thus the remainder of proof will be to show that $E_1\ge 0$.

Recall that

$$\begin{aligned}\boldsymbol{X}&=\boldsymbol{\theta}+\boldsymbol{B}\boldsymbol{Z}_0,\\\boldsymbol{Y}_i&=\boldsymbol{B}\boldsymbol{Z}_i,\quad i=1,\ldots,n,\end{aligned}$$

where $\boldsymbol{Z}_0,\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_n$ are mutually and independently distributed as $\mathcal{N}_r(\boldsymbol{0}_r,\boldsymbol{I}_r)$. Making the change of variables $\boldsymbol{U}=\boldsymbol{H}^t\boldsymbol{X}$ yields that $\boldsymbol{U}\sim\mathcal{N}_r(\boldsymbol{\xi},\boldsymbol{\Omega})$ with $\boldsymbol{\xi}=\boldsymbol{H}^t\boldsymbol{\theta}$ and $\boldsymbol{\Omega}=\boldsymbol{H}^t\boldsymbol{\Sigma}\boldsymbol{H}=\boldsymbol{B}^t\boldsymbol{B}$. Noting that, from Lemma 2.2 (4),

$$\boldsymbol{\Sigma}^+=\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-2}\boldsymbol{B}^t=\boldsymbol{H}\boldsymbol{\Omega}^{-1}\boldsymbol{H}^t,$$

we can see that

$$\boldsymbol{X}^t\boldsymbol{\Sigma}^+\boldsymbol{\theta}=\boldsymbol{X}^t\boldsymbol{H}\boldsymbol{\Omega}^{-1}\boldsymbol{H}^t\boldsymbol{\theta}=\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}.$$

Hence E_1 is expressed as

$$E_1=E[\{g_+(\boldsymbol{U})-g(\boldsymbol{U})\}\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}],$$

where $g(\boldsymbol{U})=g(\boldsymbol{H}\boldsymbol{U}+\boldsymbol{B}_0\boldsymbol{B}_0^t\boldsymbol{\theta},\boldsymbol{S})$.

The conditional expectation of $\{g_+(\boldsymbol{U})-g(\boldsymbol{U})\}\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}$ given $\boldsymbol{S}=\boldsymbol{Y}^t\boldsymbol{Y}$ is written as

$$\begin{aligned}E_0&=K(\boldsymbol{\Omega})\int_{\mathbb{R}^r}\{g_+(\boldsymbol{u})-g(\boldsymbol{u})\}\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\,e^{-(1/2)(\boldsymbol{u}-\boldsymbol{\xi})^t\boldsymbol{\Omega}^{-1}(\boldsymbol{u}-\boldsymbol{\xi})}\mathrm{d}\boldsymbol{u}\\&=K(\boldsymbol{\Omega})\int_{\mathbb{R}^r}\{g_+(\boldsymbol{u})-g(\boldsymbol{u})\}\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\,e^{\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}}e^{-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{u}/2-\boldsymbol{\xi}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}/2}\mathrm{d}\boldsymbol{u},\end{aligned}\tag{2.4}$$

where $K(\boldsymbol{\Omega})$ stands for a normalizing constant. Making the transformation $\boldsymbol{u} \to -\boldsymbol{u}$, we obtain

$$E_0 = K \int_{\mathbb{R}^r} \{g_+(\boldsymbol{u}) - g(\boldsymbol{u})\}(-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\, e^{-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}})e^{-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{u}/2-\boldsymbol{\xi}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}/2}\mathrm{d}\boldsymbol{u} \tag{2.5}$$

by the assumption that $g(\boldsymbol{u}) = g(-\boldsymbol{u})$. Hence, adding each sides of (2.4) and (2.5) yields that

$$2E_0 = K(\boldsymbol{\Omega}) \int_{\mathbb{R}^r} \{g_+(\boldsymbol{u}) - g(\boldsymbol{u})\}\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\,(e^{\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}} - e^{-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}})e^{-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{u}/2-\boldsymbol{\xi}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}/2}\mathrm{d}\boldsymbol{u}.$$

It is noted that $y(e^y - e^{-y}) \geq 0$ for any real y, which verifies that

$$\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\,(e^{\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}} - e^{-\boldsymbol{u}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}}) \geq 0.$$

Since $g_+(\boldsymbol{u}) - g(\boldsymbol{u}) \geq 0$, it is seen that $E_0 \geq 0$, namely $E_1 \geq 0$. Thus the proof is complete. □

2.3 Minimax estimation

For the mean vector $\boldsymbol{\theta}$, one of natural estimators is the unbiased estimator

$$\boldsymbol{\delta}^{UB} = \boldsymbol{X}.$$

As pointed out by Khatri (1968, page 276) and Rao (1973, page 532), $\boldsymbol{\delta}^{UB}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Since $\boldsymbol{X}$ has the stochastic representation $\boldsymbol{X} = \boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{Z}_0$ with $\boldsymbol{Z}_0 \sim \mathcal{N}_r(\boldsymbol{0}_r, \boldsymbol{I}_r)$, we observe that $\boldsymbol{Z}_0 = \boldsymbol{B}^+(\boldsymbol{X} - \boldsymbol{\theta})$, so that

$$\|\boldsymbol{Z}_0\|^2 = (\boldsymbol{X} - \boldsymbol{\theta})^t(\boldsymbol{B}^t)^+\boldsymbol{B}^+(\boldsymbol{X} - \boldsymbol{\theta}) = (\boldsymbol{X} - \boldsymbol{\theta})^t\boldsymbol{\Sigma}^+(\boldsymbol{X} - \boldsymbol{\theta}).$$

The risk of $\boldsymbol{\delta}^{UB}$ is given by

$$R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E[\|\boldsymbol{Z}_0\|^2] = r.$$

Hence $\boldsymbol{\delta}^{UB}$ has the constant risk r. We here have the following theorem.

Theorem 2.1 $\boldsymbol{\delta}^{UB}$ *is minimax relative to the loss* (1.2).

Proof. In order to prove this theorem, we consider a sequence of prior distributions and show that the corresponding sequence of the Bayes risk functions tends to the risk of $\boldsymbol{\delta}^{UB}$, namely r.

Suppose that $\boldsymbol{\theta} = \boldsymbol{B}\boldsymbol{\zeta}$ with $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^t$, where $\boldsymbol{\zeta}$ is an r-dimensional random vector. For $k = 1, 2, \dots$, define the sequence of prior distributions of $\boldsymbol{\zeta}$ as $\mathcal{N}_r(\boldsymbol{0}_r, k\boldsymbol{I}_r)$. Assume that $\boldsymbol{B}$ has a proper density proportional to

$$\pi(\boldsymbol{B}) \propto |\boldsymbol{B}^t\boldsymbol{B}|^{-p}\exp\Big\{-\frac{1}{2}\mathrm{tr}\,(\boldsymbol{B}^t\boldsymbol{B})^{-1}\Big\}.$$

The joint posterior density of $\boldsymbol{\zeta}$ and $\boldsymbol{B}$ given $\boldsymbol{X}$ and $\boldsymbol{S}=\boldsymbol{Y}^t\boldsymbol{Y}$ is proportional to

$$|\boldsymbol{B}^t\boldsymbol{B}|^{-p-(n+1)/2}\exp\Big\{-\frac{1}{2}\Big[\|\boldsymbol{B}^+(\boldsymbol{X}-\boldsymbol{B}\boldsymbol{\zeta})\|^2+\mathrm{tr}\,(\boldsymbol{B}^+)^t\boldsymbol{B}^+\boldsymbol{S}+\frac{1}{k}\|\boldsymbol{\zeta}\|^2+\mathrm{tr}\,(\boldsymbol{B}^t\boldsymbol{B})^{-1}\Big]\Big\}. \quad (2.6)$$

It is noted that $\boldsymbol{B}^+(\boldsymbol{X}-\boldsymbol{B}\boldsymbol{\zeta})=\boldsymbol{B}^+\boldsymbol{X}-\boldsymbol{\zeta}$, which yields that

$$\|\boldsymbol{B}^+(\boldsymbol{X}-\boldsymbol{B}\boldsymbol{\zeta})\|^2+\frac{1}{k}\|\boldsymbol{\zeta}\|^2=\frac{1+k}{k}\Big\|\boldsymbol{\zeta}-\frac{k}{1+k}\boldsymbol{B}^+\boldsymbol{X}\Big\|^2+\frac{1}{1+k}\boldsymbol{X}^t(\boldsymbol{B}^t)^+\boldsymbol{B}^+\boldsymbol{X}.$$

Since $\mathrm{tr}\,(\boldsymbol{B}^t\boldsymbol{B})^{-1}=\mathrm{tr}\,\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-2}\boldsymbol{B}^t=\mathrm{tr}\,(\boldsymbol{B}^+)^t\boldsymbol{B}^+$, (2.6) is rewritten as

$$|\boldsymbol{B}^t\boldsymbol{B}|^{-p-(n+1)/2}\exp\Big\{-\frac{1}{2}\Big[\frac{1+k}{k}\Big\|\boldsymbol{\zeta}-\frac{k}{1+k}\boldsymbol{B}^+\boldsymbol{X}\Big\|^2+\mathrm{tr}\,(\boldsymbol{B}^+)^t\boldsymbol{B}^+\boldsymbol{G}_k\Big]\Big\},$$

where $\boldsymbol{G}_k=\boldsymbol{I}_p+(1+k)^{-1}\boldsymbol{X}\boldsymbol{X}^t+\boldsymbol{S}$.

For each k, the resulting Bayes estimator relative to the loss (1.2) is denoted by $\boldsymbol{\delta}_k^\pi$, which must satisfy

$$E_k^\pi[\boldsymbol{\Sigma}^+(\boldsymbol{\delta}_k^\pi-\boldsymbol{\theta})]=E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]\boldsymbol{\delta}_k^\pi-E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{\zeta}]=\boldsymbol{0}_p,$$

where E_k^π indicates the posterior expectation for each k. Here $E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]$ is given by

$$E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]=K(\boldsymbol{G}_k)\int_{\mathbb{R}^{p\times r}}(\boldsymbol{B}^+)^t\boldsymbol{B}^+|\boldsymbol{B}^t\boldsymbol{B}|^{-p-(n+1)/2}\exp\Big\{-\frac{1}{2}\mathrm{tr}\,(\boldsymbol{B}^+)^t\boldsymbol{B}^+\boldsymbol{G}_k\Big\}\mathrm{d}\boldsymbol{B},$$

where $K(\boldsymbol{G}_k)$ is a normalizing constant. From Lemma 2.5, the Jacobian of transformation $\boldsymbol{C}=\boldsymbol{B}^+=(\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t$ is given by $J[\boldsymbol{B}\to\boldsymbol{C}]=|\boldsymbol{C}\boldsymbol{C}^t|^{-p}$, so that

$$E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]=K(\boldsymbol{G}_k)\int_{\mathbb{R}^{r\times p}}\boldsymbol{C}^t\boldsymbol{C}|\boldsymbol{C}\boldsymbol{C}^t|^{(n+1)/2}\exp\Big\{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{C}^t\boldsymbol{C}\boldsymbol{G}_k\Big\}\mathrm{d}\boldsymbol{C}.$$

Denoting a maximum eigenvalue of $\boldsymbol{G}_k$ by ℓ_k, we observe that

$$E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]\ge K(\boldsymbol{G}_k)\int_{\mathbb{R}^{r\times p}}\boldsymbol{C}^t\boldsymbol{C}|\boldsymbol{C}\boldsymbol{C}^t|^{(n+1)/2}\exp\Big\{-\frac{\ell_k}{2}\mathrm{tr}\,\boldsymbol{C}^t\boldsymbol{C}\Big\}\mathrm{d}\boldsymbol{C}=I_0,\quad \text{say},$$

where, for a symmetric matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, $\boldsymbol{A}_1\ge\boldsymbol{A}_2$ means $\boldsymbol{A}_1-\boldsymbol{A}_2$ is positive semi-definite. For every $\boldsymbol{O}\in\mathcal{O}(p)$, making the transformation $\boldsymbol{C}\to\boldsymbol{C}\boldsymbol{O}$ yields that $I_0=\boldsymbol{O}^tI_0\boldsymbol{O}$, which implies that I_0 has the form $c\boldsymbol{I}_p$ with $c>0$. Thus $E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]$ is positive definite for all k, so that the inverse of $E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]$ exists. Then for each k, $\boldsymbol{\delta}_k^\pi$ can be written as

$$\begin{aligned}\boldsymbol{\delta}_k^\pi&=\big\{E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]\big\}^{-1}E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{\zeta}]\\&=\frac{k}{1+k}\big\{E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]\big\}^{-1}E_k^\pi[(\boldsymbol{B}^+)^t\boldsymbol{B}^+]\boldsymbol{X}=\frac{k}{1+k}\boldsymbol{X}.\end{aligned}$$

The risk of $\boldsymbol{\delta}_k^\pi=\{k/(1+k)\}\boldsymbol{X}$ is given by

$$R(\boldsymbol{\delta}_k^\pi,\boldsymbol{\theta}|\boldsymbol{\Sigma})=\frac{k^2r}{(1+k)^2}+\frac{1}{(1+k)^2}\boldsymbol{\theta}^t\boldsymbol{\Sigma}^+\boldsymbol{\theta}=\frac{k^2r}{(1+k)^2}+\frac{1}{(1+k)^2}\|\boldsymbol{\zeta}\|^2,$$

so the Bayes risk of $\boldsymbol{\delta}_k^\pi$ is expressed as $kr/(1+k)$, which converges to r as $k\to\infty$. Hence the proof is complete. □

3 Classes of Minimax and Shrinkage Estimators

Consider the Baranchik (1970) type class of shrinkage estimators

$$\boldsymbol{\delta}^{SH} = \Big(1 - \frac{\phi(F)}{F}\Big)\boldsymbol{X},$$

where ϕ is a bounded and differentiable function of $F = \boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}$. This includes the James-Stein type shrinkage estimator

$$\boldsymbol{\delta}^{JS} = \Big(1 - \frac{a}{F}\Big)\boldsymbol{X},$$

where a is a positive constant.

Theorem 3.1 *The statistic $F = \boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}$ has the same distribution as $F = \|\boldsymbol{U}\|^2/T$, where $\boldsymbol{U}$ and T are mutually and independently distributed as $\boldsymbol{U} \sim \mathcal{N}_r(\boldsymbol{\zeta}, \boldsymbol{I}_r)$ and $T \sim \chi^2_m$ for $\boldsymbol{\zeta} = \boldsymbol{B}^+\boldsymbol{\theta}$ and $m = n - r + 1$. When the risk difference of the estimators $\boldsymbol{\delta}^{SH}$ and $\boldsymbol{\delta}^{UB}$ is denoted by*

$$\Delta = R(\boldsymbol{\delta}^{SH}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}),$$

one gets the expression $\Delta = E[\widehat{\Delta}(F)]$, where

$$\widehat{\Delta}(F) = \frac{\{(m+2)\phi(F) - 2(r-2)\}\phi(F)}{F} - 4\phi'(F)\{\phi(F) + 1\}. \tag{3.1}$$

This shows that $\widehat{\Delta}(F)$ is an unbiased estimator of the risk difference Δ.

Theorem 3.2 *The risk difference Δ given in Theorem 3.1 is rewritten as*

$$\Delta = E\Big[\frac{T}{2}F^{m/2}\int_F^\infty \frac{1}{t^{m/2+1}}\widehat{\Delta}(t)\mathrm{d}t\Big], \tag{3.2}$$

where $\widehat{\Delta}(\cdot)$ is given in (3.1). Further, one gets the expression

$$\frac{F^{m/2+1}}{2}\int_F^\infty \frac{1}{t^{m/2+1}}\widehat{\Delta}(t)\mathrm{d}t = \mathcal{I}(F), \tag{3.3}$$

where $\mathcal{I}(F)$ is defined by

$$\mathcal{I}(F) = \phi(F)\{\phi(F) + 2\} - (m + r)\int_0^1 z^{m/2}\phi(F/z)\mathrm{d}z. \tag{3.4}$$

Combining (3.2) and (3.3) shows that

$$\Delta = E\Big[\frac{T}{F}\mathcal{I}(F)\Big]. \tag{3.5}$$

Theorem 3.2 implies that the estimator $\boldsymbol{\delta}^{SH}$ is minimax if the following integral inequality holds:

$$\mathcal{I}(F) \le 0. \tag{3.6}$$

The integral inequality was first derived by Kubokawa (2009) with a slightly different way. Using the same technique as in Kubokawa (2009), Theorem 3.2 derives the integral inequality directly from $\widehat{\Delta}(F)$.

The integral inequality $\mathcal{I}(F) \le 0$ is expressed as

$$\phi(F) \le -2 + (m+r)\frac{\int_0^1 z^{m/2}\phi(F/z)\mathrm{d}z}{\phi(F)}, \tag{3.7}$$

so that one gets another sufficient condition for the minimaxity by evaluating the r.h.s. of (3.7). For instance, if $F^c\phi(F)$ is nondecreasing in F for a nonnegative constant c, then we have $(F/z)^c\phi(F/z) \ge F^c\phi(F)$ for $0 < z < 1$, so that

$$\begin{aligned}
\int_0^1 z^{m/2}\phi(F/z)\mathrm{d}z &= \int_0^1 z^{m/2}(z/F)^c\{(F/z)^c\phi(F/z)\}\mathrm{d}z \\
&\ge \int_0^1 z^{m/2}(z/F)^c\{F^c\phi(F)\}\mathrm{d}z \\
&= \int_0^1 z^{m/2+c}\mathrm{d}z\phi(F) = \frac{2}{m+2+2c}\phi(F),
\end{aligned}$$

Thus, the integral inequality (3.7) gives a simple condition $\phi(F) \le 2(r-2-2c)/(m+2+2c)$.

Proposition 3.1 *Assume that $F^c\phi(F)$ is nondecreasing in F for $c \ge 0$. If*

$$0 < \phi(F) \le \frac{2(r-2-2c)}{n-r+3+2c}, \quad \min(n,p) \ge r \ge 3, \tag{3.8}$$

then $\boldsymbol{\delta}^{SH}$ is minimax relative to the loss (1.2).

Corollary 3.1 *If*

$$0 < a \le \frac{2(r-2)}{n-r+3}, \quad \min(n,p) \ge r \ge 3,$$

then $\boldsymbol{\delta}^{JS}$ is minimax relative to the loss (1.2).

A well-known condition for the minimaxity is the differential inequality

$$\widehat{\Delta}(F) \le 0. \tag{3.9}$$

It is interesting to note that from Theorem 3.2, the differential inequality implies the integral inequality, namely, condition (3.9) is more restrictive than (3.6). For instance, we shall derive a similar condition to (3.8) from the differential inequality (3.9) under the condition that $F^c\phi(F)$ is nondecreasing in F for $c \ge 0$. It is noted that $\widehat{\Delta}(F)$ can be rewritten as

$$\widehat{\Delta}(F) = \frac{(m+2+4c)\phi^2(F) - 2(r-2-2c)\phi(F)}{F} - 4\frac{(F^c\phi(F))'}{F^c}\{\phi(F)+1\},$$

which provides a sufficient condition that (1) $F^c\phi(F)$ is nondecreasing in F for $c \ge 0$ and (2) $\phi(F) \le 2(r-2-2c)/(n-r+3+4c)$ for $m = n-r+1$. The condition (2) is more restrictive than condition (3.8) in Proposition 3.1.

Theorems 3.1 and 3.2 show that the risk of $\boldsymbol{\delta}^{SH}$ relative to the loss (1.2) is represented by expectation of a function of F. The statistic F has a noncentral F like distribution depending on the parameter $\|\boldsymbol{\zeta}\|^2$. Hence the risk of $\boldsymbol{\delta}^{SH}$ is a function of

$$\|\boldsymbol{\zeta}\|^2 = \boldsymbol{\theta}^t(\boldsymbol{B}^+)^t\boldsymbol{B}^+\boldsymbol{\theta} = \boldsymbol{\theta}^t\boldsymbol{\Sigma}^+\boldsymbol{\theta}.$$

We now give proofs of Theorems 3.1 and 3.2. The following lemmas are useful for the purpose.

Lemma 3.1 *Let $\boldsymbol{U}$ be a random vector such that $\boldsymbol{U} \sim \mathcal{N}_r(\boldsymbol{\xi}, \boldsymbol{\Omega})$. Denote by $\nabla = (\partial/\partial U_i)$ the differential operator vector with respect to $\boldsymbol{U} = (U_i)$. Let $\boldsymbol{G} = (G_i)$ be an r-dimensional function of $\boldsymbol{U}$, such that $E[|U_iG_i|] < \infty$ and $E[|\partial G_i/\partial U_j|] < \infty$ for $i, j = 1, \ldots, r$. Then we have*

$$E[(\boldsymbol{U}-\boldsymbol{\xi})^t\boldsymbol{\Omega}^{-1}\boldsymbol{G}] = E[\mathrm{tr}\,\nabla\boldsymbol{G}^t].$$

Lemma 3.2 *Let T be a random variable such that $T \sim \chi^2_m$. Let $g(t)$ be a differentiable function such that $E[|g(T)|] < \infty$ and $E[|Tg'(T)|] < \infty$. Then we have $E[Tg(T)] = E[mg(T) + 2Tg'(T)]$ and*

$$E[T^2g(T)] = E[(m+2)Tg(T) + 2T^2g'(T)].$$

Proof of Theorem 3.1. Let $\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}$. Using the same arguments as in the proof of Proposition 2.2, we observe that $\boldsymbol{U} = \boldsymbol{H}^t\boldsymbol{X} \sim \mathcal{N}_r(\boldsymbol{\xi}, \boldsymbol{\Omega})$ with $\boldsymbol{\xi} = \boldsymbol{H}^t\boldsymbol{\theta}$ and $\boldsymbol{\Omega} = \boldsymbol{B}^t\boldsymbol{B}$. Also, it follows that $\boldsymbol{S} = \boldsymbol{B}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{B}^t = \boldsymbol{H}\boldsymbol{W}\boldsymbol{H}^t$, where $\boldsymbol{W} = (\boldsymbol{B}^t\boldsymbol{B})^{1/2}\boldsymbol{Z}^t\boldsymbol{Z}(\boldsymbol{B}^t\boldsymbol{B})^{1/2} \sim \mathcal{W}_r(n, \boldsymbol{\Omega})$ independent of $\boldsymbol{U}$. Since $\boldsymbol{\Sigma}^+ = \boldsymbol{H}\boldsymbol{\Omega}^{-1}\boldsymbol{H}^t$, $\boldsymbol{H} \in \mathcal{V}_{p,r}$ and

$$\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X} = \boldsymbol{X}^t(\boldsymbol{H}\boldsymbol{W}\boldsymbol{H}^t)^+\boldsymbol{X} = \boldsymbol{X}^t(\boldsymbol{H}^t)^+\boldsymbol{W}^{-1}\boldsymbol{H}^+\boldsymbol{X} = \boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U},$$

the difference in risk between $\boldsymbol{\delta}^{SH}$ and $\boldsymbol{\delta}^{UB}$ is given by

$$\begin{aligned} R(\boldsymbol{\delta}^{SH}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) &= E\left[\frac{\phi^2(F)}{F^2}\boldsymbol{X}^t\boldsymbol{\Sigma}^+\boldsymbol{X} - 2\frac{\phi(F)}{F}(\boldsymbol{X}-\boldsymbol{\theta})^t\boldsymbol{\Sigma}^+\boldsymbol{X}\right] \\ &= E\left[\frac{\phi^2(F_1)}{F_1^2}\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{U} - 2\frac{\phi(F_1)}{F_1}(\boldsymbol{U}-\boldsymbol{\xi})^t\boldsymbol{\Omega}^{-1}\boldsymbol{U}\right] \qquad (3.10) \end{aligned}$$

with $F_1 = \boldsymbol{U}^t \boldsymbol{W}^{-1} \boldsymbol{U}$. Applying Lemma 3.1 to the second term in the last r.h.s. of (3.10) gives that

$$R(\boldsymbol{\delta}^{SH}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E\Big[\frac{\phi^2(F_1)}{F_1^2} \boldsymbol{U}^t \boldsymbol{\Omega}^{-1} \boldsymbol{U} - 2(r-2)\frac{\phi(F_1)}{F_1} - 4\phi'(F_1)\Big]. \quad (3.11)$$

It is a well-known fact that $\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{U}/\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U} \sim \chi^2_{n-r+1}$ independent of $\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{U}$. Letting $U = \boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{U}$ and $T = \boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{U}/\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U}$, we obtain $F_1 = U/T$ and rewrite (3.11) as

$$R(\boldsymbol{\delta}^{SH}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E\Big[\frac{T^2\phi^2(U/T)}{U} - 2(r-2)\frac{T\phi(U/T)}{U} - 4\phi'\Big(\frac{U}{T}\Big)\Big].$$

Applying Lemma 3.2 to the first term of the r.h.s. yields that

$$E\Big[\frac{T^2\phi^2(U/T)}{U}\Big] = E^U\Big[E^{T|U}\Big[\frac{T^2\phi^2(U/T)}{U}\Big]\Big] = E\Big[(n-r+3)\frac{T\phi^2(U/T)}{U} - 4\phi\Big(\frac{U}{T}\Big)\phi'\Big(\frac{U}{T}\Big)\Big],$$

so that

$$\begin{aligned}
&R(\boldsymbol{\delta}^{SH}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) \\
&= E\Big[\Big\{(n-r+3)\phi\Big(\frac{U}{T}\Big) - 2(r-2)\Big\}\frac{T\phi(U/T)}{U} - 4\phi'\Big(\frac{U}{T}\Big)\Big\{\phi\Big(\frac{U}{T}\Big)+1\Big\}\Big] \\
&= E[\{(n-r+3)\phi(F) - 2(r-2)\}\phi(F)/F - 4\phi'(F)\{\phi(F)+1\}].
\end{aligned}$$

Making the transformation $\boldsymbol{\Omega}^{-1/2}\boldsymbol{U} \to \boldsymbol{U}$ gives that the resulting $\boldsymbol{U}$ is distributed as $\mathcal{N}_r(\boldsymbol{\zeta}, \boldsymbol{I}_r)$ with $\boldsymbol{\zeta} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\xi} = (\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t\boldsymbol{\theta} = \boldsymbol{B}^+\boldsymbol{\theta}$. Hence the proof is complete. □

Proof of Theorem 3.2. Let $U = \|\boldsymbol{U}\|^2$. For $F = U/T$, let

$$G(F) = \frac{F^{m/2}}{2}\int_F^\infty \frac{1}{t^{m/2+1}}\widehat{\Delta}(t)\mathrm{d}t.$$

Then from Lemma 3.2, it follows that

$$E^{T|U}[TG(F)] = E^{T|U}[mG(F) - 2T\frac{U}{T^2}G'(F)],$$

where $E^{T|U}[\cdot]$ denotes the conditional expectation with respect to T given U. Since $mG(F) - 2FG'(F) = \widehat{\Delta}(F)$, it is easy to see that

$$E[TG(F)] = E[\widehat{\Delta}(F)],$$

which shows (3.2). To show the equality (3.3), it is noted that

$$\int_F^\infty \frac{\widehat{\Delta}(t)}{t^{m/2+1}}\mathrm{d}t = \int_F^\infty \Big\{\frac{(m+2)\phi^2(t) - 2(r-2)\phi(t)}{t^{m/2+2}} - 4\frac{\phi'(t)\{\phi(t)+1\}}{t^{m/2+1}}\Big\}\mathrm{d}t. \quad (3.12)$$

By integration by parts, it is observed that

$$\int_F^\infty \frac{\phi'(t)}{t^{m/2+1}}\mathrm{d}t = -\frac{\phi(F)}{F^{m/2+1}} + \frac{m+2}{2}\int_F^\infty \frac{\phi(t)}{t^{m/2+2}}\mathrm{d}t,$$

$$\int_F^\infty \frac{\phi(t)\phi'(t)}{t^{m/2+1}}\mathrm{d}t = -\frac{1}{2}\frac{\phi^2(F)}{F^{m/2+1}} + \frac{m+2}{4}\int_F^\infty \frac{\phi^2(t)}{t^{m/2+2}}\mathrm{d}t.$$

Then from (3.12), we have

$$\begin{aligned}\int_F^\infty \frac{\widehat{\Delta}(t)}{t^{m/2+1}}\mathrm{d}t =& -2(m+r)\int_F^\infty \frac{\phi(t)}{t^{m/2+2}}\mathrm{d}t + 2\frac{\phi^2(F)}{F^{m/2+1}} + 4\frac{\phi(F)}{F^{m/2+1}} \\ =&\frac{2}{F^{m/2+1}}\Big\{\phi(F)\{\phi(F)+2\} - (m+r)\int_0^1 z^{m/2}\phi(F/z)\mathrm{d}z\Big\}, \end{aligned} \tag{3.13}$$

where the second equality is derived by making the transformation $z = F/t$ with $\mathrm{d}z = -(F/t^2)\mathrm{d}t$. The expression given in (3.13) shows (3.3) and (3.4). Finally, the equality (3.5) follows since $\Delta = E[TG(F)] = E[(T/F)\mathcal{I}(F)]$. □

4 A Bayesian Motivation and Generalized Bayes and Minimax Estimators

4.1 Empirical Bayes approach

We begin by giving a Bayesian motivation for the James-Stein type shrinkage estimator. Under the singular case of the covariance matrix, we here demonstrate that the James-Stein (1961) type shrinkage estimator can be provided as an empirical Bayes estimator using an argument similar to that in Efron and Morris (1972).

The empirical Bayesian approach considered here consists of the following steps: (1) Reduce the estimation problem to that on the r-dimensional subspace in $\mathbb{R}^p$ spanned by column vectors of $\boldsymbol{B}$; (2) Derive an empirical Bayes estimator on the subspace; (3) Return the Bayes estimator to the original whole space.

Step (1). Let

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2},$$

where $(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}$ is a symmetric square root of $(\boldsymbol{B}^t\boldsymbol{B})^{-1}$. Note that $\boldsymbol{H} \in \mathcal{V}_{p,r}$ and

$$\boldsymbol{H}\boldsymbol{H}^t = \Sigma\Sigma^+ = \boldsymbol{S}\boldsymbol{S}^+.$$

Since $\boldsymbol{X} = \boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{Z}_0$ with $\boldsymbol{Z}_0 \sim \mathcal{N}_r(\boldsymbol{0}_r, \boldsymbol{I}_r)$, it follows that

$$\boldsymbol{H}^t\boldsymbol{X} = \boldsymbol{H}^t\boldsymbol{\theta} + \boldsymbol{H}^t\boldsymbol{B}\boldsymbol{Z}_0 = \boldsymbol{H}^t\boldsymbol{\theta} + (\boldsymbol{B}^t\boldsymbol{B})^{1/2}\boldsymbol{Z}_0,$$

so that $\boldsymbol{H}^t\boldsymbol{X} \sim \mathcal{N}_r(\boldsymbol{\xi}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \boldsymbol{B}^t\boldsymbol{B}$ and

$$\boldsymbol{\xi} = \boldsymbol{H}^t\boldsymbol{\theta}. \tag{4.1}$$

For the stochastic representation $\boldsymbol{S} = \boldsymbol{B}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{B}^t$, $\boldsymbol{S}$ can be expressed by

$$\boldsymbol{S} = \boldsymbol{H}\boldsymbol{W}\boldsymbol{H}^t,$$

where $\boldsymbol{W} = (\boldsymbol{B}^t\boldsymbol{B})^{1/2}\boldsymbol{Z}^t\boldsymbol{Z}(\boldsymbol{B}^t\boldsymbol{B})^{1/2} \sim \mathcal{W}_r(n, \boldsymbol{B}^t\boldsymbol{B})$. Also, it is seen that $\boldsymbol{H}^t\boldsymbol{S}\boldsymbol{H} = \boldsymbol{W}$ and

$$\boldsymbol{\Sigma}^+ = \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-2}\boldsymbol{B}^t = \boldsymbol{H}\boldsymbol{\Omega}^{-1}\boldsymbol{H}^t. \tag{4.2}$$

Then we assume that $\boldsymbol{H}$ is fixed and rewrite the likelihood of $\boldsymbol{X}$ and $\boldsymbol{S}$ as

$$\begin{aligned} &L(\boldsymbol{U}, \boldsymbol{W}|\boldsymbol{\xi}, \boldsymbol{\Omega}) \\ &= K|\boldsymbol{\Omega}|^{-(n+1)/2}|\boldsymbol{W}|^{(n-r-1)/2}\exp\Big\{-\frac{1}{2}(\boldsymbol{U}-\boldsymbol{\xi})^t\boldsymbol{\Omega}^{-1}(\boldsymbol{U}-\boldsymbol{\xi}) - \frac{1}{2}\mathrm{tr}\,\boldsymbol{\Omega}^{-1}\boldsymbol{W}\Big\}, \end{aligned} \tag{4.3}$$

where $\boldsymbol{U} = \boldsymbol{H}^t\boldsymbol{X}$ and K is a positive constant. It is noted that $\boldsymbol{U}$ and $\boldsymbol{W}$ are random and $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ are unknown parameters.

Step (2). We here consider the following prior distribution for $\boldsymbol{\xi} = \boldsymbol{H}^t\boldsymbol{\theta}$:

$$\boldsymbol{\xi} \sim \mathcal{N}_r(\boldsymbol{0}_r, \{(1-\eta)/\eta\}\boldsymbol{\Omega}), \quad 0 < \eta < 1, \tag{4.4}$$

where η is a hyperparameter. Multiplying (4.3) by the density of (4.4), we observe that the posterior distribution of $\boldsymbol{\xi}$ given $\boldsymbol{U}$ and $\boldsymbol{W}$ is

$$\boldsymbol{\xi}|\boldsymbol{U}, \boldsymbol{W} \sim \mathcal{N}_r\big(\widehat{\boldsymbol{\xi}}^B(\boldsymbol{U}, \boldsymbol{W}, \eta), (1-\eta)\boldsymbol{\Omega}\big),$$

where $\widehat{\boldsymbol{\xi}}^B(\boldsymbol{U}, \boldsymbol{W}, \eta) = (1-\eta)\boldsymbol{U}$. Also, the marginal density is proportional to

$$\eta^{r/2}|\boldsymbol{\Omega}|^{-(n+1)/2}|\boldsymbol{W}|^{(n-r-1)/2}\exp\Big[-\frac{\eta}{2}\boldsymbol{U}^t\boldsymbol{\Omega}^{-1}\boldsymbol{U} - \frac{1}{2}\mathrm{tr}\,\boldsymbol{\Omega}^{-1}\boldsymbol{W}\Big]. \tag{4.5}$$

Since $\widehat{\boldsymbol{\xi}}^B(\boldsymbol{U}, \boldsymbol{W}, \eta)$ is the Bayes estimator relative to the squared errors loss, the Bayes estimator of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\delta}^B(\boldsymbol{U}, \boldsymbol{W}, \eta) = \boldsymbol{H}\widehat{\boldsymbol{\xi}}^B(\boldsymbol{U}, \boldsymbol{W}, \eta) = (1-\eta)\boldsymbol{H}\boldsymbol{U}.$$

The parameter η in $\boldsymbol{\delta}^B(\boldsymbol{U}, \boldsymbol{W}, \eta)$ requires to be estimated from the marginal density (4.5), which is equivalent to the model that $\boldsymbol{U} \sim \mathcal{N}_r(\boldsymbol{0}_r, \boldsymbol{\Omega}/\eta)$ and $\boldsymbol{W} \sim \mathcal{W}_r(n, \boldsymbol{\Omega})$. A reasonable estimator of η is of the form, for a positive constant a,

$$\widehat{\eta} = \frac{a}{\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U}},$$

which includes the unbiased estimator

$$\widehat{\eta}^{UB} = \frac{r-2}{(n-r+1)\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U}}.$$

The estimator $\widehat{\eta}$ is substituted into the Bayes estimator $\boldsymbol{\delta}^B(\boldsymbol{U}, \boldsymbol{W}, \eta)$ to get the empirical Bayes estimator

$$\boldsymbol{\delta}^{EB} = \Big(1 - \frac{a}{\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U}}\Big)\boldsymbol{H}\boldsymbol{U} = \Big(1 - \frac{a}{\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}}\Big)\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}$$

because $\boldsymbol{H}\boldsymbol{U} = \boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X} = \boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}$ and

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

Since η is restricted to $0 < \eta < 1$, we should find an optimal estimator subject to $0 < \eta < 1$. For instance, such optimal estimator is given by

$$\widehat{\eta}^{TR} = \min\Big\{1, \frac{a}{\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U}}\Big\} = \min\Big\{1, \frac{a}{\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}}\Big\},$$

which is used to obtain

$$\boldsymbol{\delta}^{TR} = \Big(1 - \min\Big\{1, \frac{a}{\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}}\Big\}\Big)\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} = \max\Big\{0, 1 - \frac{a}{\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}}\Big\}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}.$$

Step (3). From Proposition 2.1, $\boldsymbol{\delta}^{EB}$ has the same risk as the James-Stein type shrinkage estimator

$$\boldsymbol{\delta}_1^{JS} = \Big(1 - \frac{a}{\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}}\Big)\boldsymbol{X}.$$

Similarly, $\boldsymbol{\delta}^{TR}$ has the same risk as

$$\boldsymbol{\delta}_1^{TR} = \max\Big\{0, 1 - \frac{a}{\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}}\Big\}\boldsymbol{X},$$

which is a positive-part James-Stein type shrinkage estimator. Thus the James-Stein type and the positive part James-Stein type shrinkage estimators can be characterized by empirical Bayes approach.

4.2 Generalized Bayes and minimax estimators

In this subsection we derive generalized Bayes estimators for a hierarchical prior and discuss the minimaxity. The hierarchical prior is analogous to that of Lin and Tsai (1973) in a nonsingular multivariate normal model.

A generalized Bayes estimator $\boldsymbol{\delta}_0^{GB}$ relative to the loss (1.2) needs to satisfy

$$E^\pi[\boldsymbol{\Sigma}^+(\boldsymbol{\delta}_0^{GB} - \boldsymbol{\theta})] = \boldsymbol{0}_p, \tag{4.6}$$

where E^π denotes the posterior expectation given $\boldsymbol{X}$ and $\boldsymbol{Y}$. Now, using (4.1) and (4.2) with the assumption that $\boldsymbol{H}$ is fixed, we rewrite the expression (4.6) as

$$\begin{aligned}
E^\pi[\boldsymbol{\Sigma}^+(\boldsymbol{\delta}_0^{GB} - \boldsymbol{\theta})] &= E^\pi[\boldsymbol{\Sigma}^+]\boldsymbol{\delta}_0^{GB} - E^\pi[\boldsymbol{\Sigma}^+\boldsymbol{\theta}] \\
&= \boldsymbol{H}E^\pi[\boldsymbol{\Omega}^{-1}]\boldsymbol{H}^t\boldsymbol{\delta}_0^{GB} - \boldsymbol{H}E^\pi[\boldsymbol{\Omega}^{-1}\boldsymbol{H}^t\boldsymbol{\theta}] \\
&= \boldsymbol{H}E^\pi[\boldsymbol{\Omega}^{-1}]\boldsymbol{H}^t\boldsymbol{\delta}_0^{GB} - \boldsymbol{H}E^\pi[\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}] = \boldsymbol{0}_p,
\end{aligned}$$

so that

$$\boldsymbol{H}^t\boldsymbol{\delta}_0^{GB} = \big\{E^\pi[\boldsymbol{\Omega}^{-1}]\big\}^{-1}E^\pi[\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}]. \tag{4.7}$$

Here, we define a prior distribution of $(\boldsymbol{\xi},\boldsymbol{\Omega})$. Assume that $(\boldsymbol{\xi},\boldsymbol{\Omega})$ has a hierarchical prior distribution and the joint density is given by

$$\pi(\boldsymbol{\xi},\boldsymbol{\Omega}) \propto \int_0^1 \pi(\boldsymbol{\xi},\boldsymbol{\Omega}|\eta)\pi(\eta)\mathrm{d}\eta, \tag{4.8}$$

where

$$\pi(\boldsymbol{\xi},\boldsymbol{\Omega}|\eta) \propto |\boldsymbol{\Omega}|^{-b/2}\Big|\frac{1-\eta}{\eta}\boldsymbol{\Omega}\Big|^{-1/2}\exp\Big\{-\frac{\eta}{2(1-\eta)}\boldsymbol{\xi}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\Big\}, \qquad \pi(\eta)\propto\eta^{a/2}$$

with constants a and b. Multiplying (4.3) by (4.8) yields the joint posterior density of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ given $\boldsymbol{X}$ and $\boldsymbol{Y}$, which is proportional to

$$\begin{aligned}\int_0^1 &|(1-\eta)\boldsymbol{\Omega}|^{-1/2}\exp\Big[-\frac{1}{2(1-\eta)}\{\boldsymbol{\xi}-(1-\eta)\boldsymbol{U}\}^t\boldsymbol{\Omega}^{-1}\{\boldsymbol{\xi}-(1-\eta)\boldsymbol{U}\}\Big]\\ &\times|\boldsymbol{\Omega}|^{-(n+b+1)/2}\exp\Big[-\frac{1}{2}\mathrm{tr}\,\boldsymbol{\Omega}^{-1}(\eta\boldsymbol{U}\boldsymbol{U}^t+\boldsymbol{W})\Big]\times\eta^{(a+r)/2}\mathrm{d}\eta\end{aligned}$$

for $a+r>-2$ and $n+b-2r\geq 0$. Therefore, integrating out (4.7) with respect to $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ gives that

$$\boldsymbol{H}^t\boldsymbol{\delta}_0^{GB} = \{E_\eta^\pi[(\eta\boldsymbol{U}\boldsymbol{U}^t+\boldsymbol{W})^{-1}]\}^{-1}E_\eta^\pi[(1-\eta)(\eta\boldsymbol{U}\boldsymbol{U}^t+\boldsymbol{W})^{-1}]\boldsymbol{U},$$

where E_η^π stands for expectation with respect to the posterior density of η given $\boldsymbol{U}$ and $\boldsymbol{W}$, which is proportional to

$$\begin{aligned}\pi(\eta|\boldsymbol{U},\boldsymbol{W}) &\propto \eta^{(a+r)/2}|\eta\boldsymbol{U}\boldsymbol{U}^t+\boldsymbol{W}|^{-(n+b-r)/2}\\ &\propto \eta^{(a+r)/2}(1+\eta F)^{-(n+b-r)/2}\end{aligned}$$

with $F=\boldsymbol{U}^t\boldsymbol{W}^{-1}\boldsymbol{U}$. A simple manipulation leads to

$$\begin{aligned}\boldsymbol{H}^t\boldsymbol{\delta}_0^{GB} &= \Big\{1-\frac{E_\eta^\pi[\eta(1+\eta F)^{-1}]}{E_\eta^\pi[(1+\eta F)^{-1}]}\Big\}\boldsymbol{U}\\ &= \Big\{1-\frac{\int_0^1\eta\cdot\eta^{(a+r)/2}(1+\eta F)^{-(n+b-r+2)/2}\mathrm{d}\eta}{\int_0^1\eta^{(a+r)/2}(1+\eta F)^{-(n+b-r+2)/2}\mathrm{d}\eta}\Big\}\boldsymbol{U},\end{aligned}$$

which yields that

$$\boldsymbol{H}^t\boldsymbol{\delta}_0^{GB} = \Big(1-\frac{\phi^{GB}(F)}{F}\Big)\boldsymbol{H}^t\boldsymbol{X}, \tag{4.9}$$

where $F=\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}$ and

$$\phi^{GB}(F) = \frac{\int_0^F\eta\cdot\eta^{(a+r)/2}(1+\eta)^{-(n+b-r+2)/2}\mathrm{d}\eta}{\int_0^F\eta^{(a+r)/2}(1+\eta)^{-(n+b-r+2)/2}\mathrm{d}\eta}.$$

The expression (4.9) indicates a system of linear equations in $\boldsymbol{\delta}_0^{GB}$. Since $\boldsymbol{H} \in \mathcal{V}_{p,r}$, using Lemma 2.4 yields the reasonable solution of the system (4.9),

$$\boldsymbol{\delta}_0^{GB} = \Big(1 - \frac{\phi^{GB}(F)}{F}\Big)\boldsymbol{X} + (\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{a},$$

where $\boldsymbol{S}\boldsymbol{S}^+ = \boldsymbol{H}\boldsymbol{H}^t$ and $\boldsymbol{a}$ is an arbitrary p-dimensional vector. From Proposition 2.1, the vector $(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{a}$ has no effect on the risk of $\boldsymbol{\delta}_0^{GB}$, so that we define the resulting generalized Bayes estimator as

$$\boldsymbol{\delta}^{GB} = \Big(1 - \frac{\phi^{GB}(F)}{F}\Big)\boldsymbol{X}.$$

It is easy to verify that $\phi^{GB}(F)$ is nondecreasing in F and

$$\lim_{F\to\infty} \phi^{GB}(F) = \frac{a+r+2}{n-a+b-2r-2}.$$

Using Proposition 3.1, we obtain the following proposition.

Proposition 4.1 *Assume that $a+r > -2$, $n+b-2r \geq 0$ and*

$$0 < \frac{a+r+2}{n-a+b-2r-2} \leq \frac{2(r-2)}{n-r+3}.$$

Then $\boldsymbol{\delta}^{GB}$ is minimax relative to the loss (1.2).

In the case such that $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_p$ for an unknown positive parameter σ^2, several priors have been proposed for constructing minimax estimators. For instance, see Maruyama and Strawderman (2005), Wells and Zhou (2008) and Kubokawa (2009). Their priors can be applied to our singular case and we can derive some classes of generalized Bayes minimax estimators improving on the James-Stein type shrinkage estimator. However, the detailed discussion is omitted here.

5 Extensions and Remarks

In the previous sections, under the assumption $\min(n,p) \geq r$, we have derived classes of minimax and shrinkage estimators. Out of the classes, we have singled out the generalized Bayes and minimax estimators with hierarchical prior. In this section, we mention some extensions and remarks.

Concerning the class of minimax estimators, we can treat the case $p \geq r > n$ as well as the case $\min(n,p) \geq r$. In the case $p \geq r > n$, consider the Chételat and Wells (2012) type class of shrinkage estimators given by

$$\boldsymbol{\delta}^{CW} = \boldsymbol{X} - \frac{\phi(F)}{F}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X},$$

where $\phi(F)$ is a bounded and differentiable function of $F = \boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X}$.

Proposition 5.1 *In the case $p \geq r > n$, the risk difference of the estimators $\boldsymbol{\delta}^{CW}$ and $\boldsymbol{\delta}^{UB}$ is written as $R(\boldsymbol{\delta}^{CW}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E[\widehat{\Delta}(F)]$, where $\widehat{\Delta}(F)$ is given in* (3.1) *for $m = r - n + 1$ and $F = \boldsymbol{X}^t \boldsymbol{S}^+ \boldsymbol{X}$.*

Proof. In the case such that $p > r > n$, the rank of $\boldsymbol{\Sigma}$ is r, while that of $\boldsymbol{S}$ is n. Write $\boldsymbol{S}$ as

$$\boldsymbol{S} = \boldsymbol{B}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{B}^t = \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{W}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t,$$

where $\boldsymbol{W} = (\boldsymbol{B}^t\boldsymbol{B})^{1/2}\boldsymbol{Z}^t\boldsymbol{Z}(\boldsymbol{B}^t\boldsymbol{B})^{1/2}$ is the $r \times r$ singular Wishart matrix with n degrees of freedom and mean $n\boldsymbol{B}^t\boldsymbol{B}$. Since $\boldsymbol{W}$ is of rank n, it can be decomposed as $\boldsymbol{W} = \boldsymbol{R}\boldsymbol{L}\boldsymbol{R}^t$, where $\boldsymbol{R} \in \mathcal{V}_{r,n}$ and $\boldsymbol{L} \in \mathcal{D}_n$. Noting that $\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{R} \in \mathcal{V}_{p,n}$, we observe from Lemma 2.3 that

$$\begin{aligned}
\boldsymbol{S}^+ &= \{\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{R}\boldsymbol{L}\boldsymbol{R}^t(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t\}^+ \\
&= \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{R}\boldsymbol{L}^{-1}\boldsymbol{R}^t(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t \\
&= \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{W}^+(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t,
\end{aligned}$$

which leads to

$$\boldsymbol{S}\boldsymbol{S}^+ = \boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{W}\boldsymbol{W}^+(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t.$$

Let $\boldsymbol{U} = (\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t\boldsymbol{X}$, $\boldsymbol{\xi} = (\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t\boldsymbol{\theta}$ and $\boldsymbol{\Omega} = \boldsymbol{B}^t\boldsymbol{B}$. It follows that $\boldsymbol{U} \sim \mathcal{N}_r(\boldsymbol{\xi}, \boldsymbol{\Omega})$. Note that $\boldsymbol{\Omega}$ is positive definite and

$$\boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X} = \boldsymbol{X}^t\boldsymbol{B}(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{W}^+(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}\boldsymbol{B}^t\boldsymbol{X} = \boldsymbol{U}^t\boldsymbol{W}^+\boldsymbol{U}.$$

Similarly, we obtain

$$(\boldsymbol{X} - \boldsymbol{\theta})^t\boldsymbol{\Sigma}^+\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} = (\boldsymbol{U} - \boldsymbol{\xi})^t\boldsymbol{\Omega}^{-1}\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{U}$$

and

$$\boldsymbol{X}^t\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\Sigma}^+\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} = \boldsymbol{U}^t\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{\Omega}^{-1}\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{U}.$$

Thus the difference in risk of $\boldsymbol{\delta}^{CW}$ and $\boldsymbol{\delta}^{UB}$ can be written as

$$\begin{aligned}
&R(\boldsymbol{\delta}^{CW}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) \\
&= E\left[\frac{\phi^2(F)}{F^2}\boldsymbol{X}^t\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{\Sigma}^+\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X} - 2\frac{\phi(F)}{F}(\boldsymbol{X} - \boldsymbol{\theta})^t\boldsymbol{\Sigma}^+\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}\right] \\
&= E\left[\frac{\phi^2(F_1)}{F_1^2}\boldsymbol{U}^t\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{\Omega}^{-1}\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{U} - 2\frac{\phi(F_1)}{F_1}(\boldsymbol{U} - \boldsymbol{\xi})^t\boldsymbol{\Omega}^{-1}\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{U}\right],
\end{aligned}$$

where $F_1 = \boldsymbol{U}^t\boldsymbol{W}^+\boldsymbol{U}$. Note that $\boldsymbol{W} \sim \mathcal{W}_r(n, \boldsymbol{\Omega})$ and $\mathrm{tr}\,\boldsymbol{W}\boldsymbol{W}^+ = n < r$. Using the same arguments as in the proof of Theorem 1 in Chételat and Wells (2012), we can see that

$$\begin{aligned}
&R(\boldsymbol{\delta}^{CW}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) \\
&= E\left[\phi^2(F_1)\frac{n + r - 2\mathrm{tr}\,\boldsymbol{W}\boldsymbol{W}^+ + 3}{F_1} - 2\phi(F_1)\frac{\mathrm{tr}\,\boldsymbol{W}\boldsymbol{W}^+ - 2}{F_1} - 4\phi'(F_1)\{\phi(F_1) + 1\}\right] \\
&= E\left[\frac{\{(r - n + 3)\phi(F) - 2(n - 2)\}\phi(F)}{F} - 4\phi'(F)\{\phi(F) + 1\}\right],
\end{aligned}$$

where $F = \boldsymbol{X}^t \boldsymbol{S}^+ \boldsymbol{X}$. $\square$

In the case $p > r > n$, the minimaxity of the unbiased estimator can be established by the same Bayesian method as in Section 2.3. Also, a positive-part rule is applicable to $\boldsymbol{\delta}^{CW}$, which is verified by the same lines as in Tsukuma and Kubokawa (2014). It is, however, hard to construct a generalized Bayes minimax estimator in the case $p \geq r > n$.

Since $p \geq r$, as possible ordering among p, n and r, we can consider the three cases: $n \geq p = r$, $\min(n, p) > r$ and $p \geq r > n$. The first case is a standard nonsingular model, and the risk expression was given in the literature. The rest of the cases have been treated in Theorem 3.1 and Proposition 5.1 of this paper. Combining these results yields a unified expression of the risk difference.

Proposition 5.2 *For the three cases $n \geq p = r$, $\min(n, p) > r$ and $p \geq r > n$, consider the class of the shrinkage estimators of the form $\boldsymbol{\delta}^{CW} = \boldsymbol{X} - \{\phi(F)/F\}\boldsymbol{S}\boldsymbol{S}^+\boldsymbol{X}$ for $F = \boldsymbol{X}^t \boldsymbol{S}^+ \boldsymbol{X}$. Then the unified expression of the risk difference between $\boldsymbol{\delta}^{CW}$ and $\boldsymbol{\delta}^{UB}$ is given by $R(\boldsymbol{\delta}^{CW}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) - R(\boldsymbol{\delta}^{UB}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E[\widehat{\Delta}(F)]$ where*

$$\widehat{\Delta}(F) = \frac{\{(m+2)\phi(F) - 2(\min(n, r) - 2)\}\phi(F)}{F} - 4\phi'(F)\{\phi(F) + 1\},$$

for $m = |n - r| + 1$.

We conclude this section with some remarks on the related topics. It is noted that the loss function given in (1.2) measures the accuracy of estimators on the r-dimensional subspace in $\mathbb{R}^p$ spanned by column vectors of $\boldsymbol{B}$. We should possibly measure the accuracy on the whole space $\mathbb{R}^p$ and employ a quadratic loss with nonsingular weight matrix. Such quadratic loss is, for instance,

$$L_Q(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{B}, \boldsymbol{B}_0) = (\boldsymbol{\delta} - \boldsymbol{\theta})^t(\boldsymbol{\Sigma}^+ + \boldsymbol{B}_0\boldsymbol{B}_0^t)(\boldsymbol{\delta} - \boldsymbol{\theta}),$$

where $\boldsymbol{B}_0 \in \mathcal{V}_{p,p-r}$ such that $\boldsymbol{B}_0^+\boldsymbol{B} = \boldsymbol{B}_0^t\boldsymbol{B} = \boldsymbol{0}_{(p-r)\times r}$. It is easy to handle the risk of shrinkage type estimators

$$\boldsymbol{\delta}^S = \left\{1 - \frac{\phi(F_0)}{F_0}\right\}\boldsymbol{X}, \quad F_0 = \boldsymbol{X}^t\boldsymbol{S}^+\boldsymbol{X} + \boldsymbol{X}^t(\boldsymbol{I}_p - \boldsymbol{S}\boldsymbol{S}^+)\boldsymbol{X}.$$

Some dominance results can be provided for the above shrinkage estimators and their positive part estimators.

Finally, we give a note on invariance of models. In the nonsingular case $r = p$, namely, $n > p$, the covariance matrix $\boldsymbol{\Sigma}$ is nonsingular and the estimation problem considered in this paper is invariant under the group of transformations

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

for a $p \times p$ nonsingular matrix $\boldsymbol{P}$. The invariance is very important and verifies intuitively reasonable estimators of the mean vector $\boldsymbol{\theta}$. Using the invariance shows that the risk of the Baranchik type shrinkage estimator $\boldsymbol{\delta}^{SH}$ is a function of $\boldsymbol{\theta}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}$. In the singular case $r < p$, on the other hand, the estimation problem with the loss (1.2) does not preserve invariance since $(\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^t)^+ \neq (\boldsymbol{P}^t)^{-1}\boldsymbol{\Sigma}^+\boldsymbol{P}^{-1}$ except when $\boldsymbol{P}$ is an orthogonal matrix. However the risk of $\boldsymbol{\delta}^{SH}$ is expressed as a function of $\boldsymbol{\theta}^t\boldsymbol{\Sigma}^+\boldsymbol{\theta}$.

Acknowledgments

The work of the first author is supported by Grant-in-Aid for Scientific Research (23500361), Japan. The work of the second author is supported by Grant-in-Aid for Scientific Research (21540114 and 23243039), 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] Díaz-García, J.A., Gutierrez-Jáimez, R. and Mardia, K.V. (1997). Wishart and pseudo-Wishart distributions and some applications to shape theory, *J. Multivariate Anal.*, **63**, 73–87.

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

[5] Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution, *Ann. Statist.*, **4**, 11–21.

[6] Harville, D.A. (1997). *Matrix algebra from a statistician's perspective*, Springer, New York.

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

[8] Khatri, C.G. (1968). Some results for the singular normal multivariate regression models, *Sankhyā, Ser. A*, **30** 267–280.

[9] Kubokawa, T. (2009). Integral inequality for minimaxity in the Stein problem, *J. Japan Statist. Soc.*, **39**, 155–175.

[10] Lin, P.-E. and Tsai, H.-L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix, *Ann. Statist.*, **1**, 142–145.

[11] Maruyama, Y. and Strawderman, W.E. (2005). A new class of generalized Bayes minimax ridge regression estimators, *Ann. Statist.*, **33**, 1753–1770.

[12] Olkin, I. (1998). The density of the inverse and pseudo-inverse of a random matrix, *Statist. & Probab. Letters*, **38**, 131–135.

[13] Rao, C.R. (1973). *Linear Statistical Inference and its Applications (2nd ed.)*, Wiley, New York.

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

[15] Srivastava, M.S. and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics*, North Holland, New York.

[16] Tsukuma, H. and Kubokawa, T. (2014). A unified approach to estimating a normal mean matrix in high and low dimensions, Discussion Paper Series, CIRJE-F-926, Faculty of Economics, The University of Tokyo. This is downloadable from http://www.cirje.e.u-tokyo.ac.jp/index.html.

[17] Wells, M. and Zhou, G. (2008). Generalized Bayes minimax estimators of the mean of multivariate normal distribution with unknown variance, *J. Multivariate Anal.*, **99**, 2208–2220.

[18] Zhang, Y. (1985). The exact distribution of the Moore-Penrose inverse of X with a density, in *Multivariate Analysis VI*, P.R. Krishnaiah (Ed.), Elsevier, New York.