Ann. Inst. Statist. Math.
Vol. 53, No. 4, 760–768 (2001)


# ESTIMATION OF THE MULTIVARIATE NORMAL PRECISION MATRIX UNDER THE ENTROPY LOSS *

XIAN ZHOU[1], XIAOQIAN SUN[2] AND JINGLONG WANG[2]

[1]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China,* e-mail: maxzhou@polyu.edu.hk
[2]*Department of Statistics, East China Normal University, Shanghai 200062, China*

(Received May 12, 1999; revised November 15, 1999)

**Abstract.** Let $X_1, \dots, X_n (n > p)$ be a random sample from multivariate normal distribution $N_p(\mu, \Sigma)$, where $\mu \in R^p$ and $\Sigma$ is a positive definite matrix, both $\mu$ and $\Sigma$ being unknown. We consider the problem of estimating the precision matrix $\Sigma^{-1}$. In this paper it is shown that for the entropy loss, the best lower-triangular affine equivariant minimax estimator of $\Sigma^{-1}$ is inadmissible and an improved estimator is explicitly constructed. Note that our improved estimator is obtained from the class of lower-triangular scale equivariant estimators.

*Key words and phrases*: Best lower-triangular equivariant minimax estimator, precision matrix, inadmissibility, multivariate normal distribution, risk function, the entropy loss.

## 1. Introduction

Let $X_1, \dots, X_n$ be the independent observations from a multivariate normal distribution $N_p(\mu, \Sigma)$ where both the mean vector $\mu \in R^p$ and the covariance matrix $\Sigma > 0$ are unknown. It is well known that $(\bar{X}, S)$ is a complete sufficient statistic for $\theta = (\mu, \Sigma)$ and $\bar{X}$ is independent of $S$, where $\bar{X} = n^{-1} \sum_{i=1}^n X_i \sim N_p(\mu, \Sigma/n)$ and $S = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \sim W_p(n-1, \Sigma)$ denote the sample mean vector and the sample dispersion matrix, respectively, and $T$ denotes the transpose of a matrix or vector. For estimating the generalized variance $|\Sigma|$, the generalized precision $|\Sigma|^{-1}$ and the covariance matrix $\Sigma$, many people have obtained important results such as Kubokawa (1989), Pal (1988), Sinha and Ghosh (1987), Sugiura (1988), Sun (1998), Takemura (1984), Wang (1984), etc. Here we are interested in estimating the precision matrix $\Sigma^{-1}$.

Under the squared loss $L_1(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \mathrm{tr}(\hat{\Sigma}^{-1}\Sigma - I_p)^2$ where $I_p$ denotes the identity matrix of order $p$, Haff (1979) showed that, under the affine transformation group $\mathcal{A}$:

$$(\bar{X}, S) \to (A\bar{X} + B, ASA^T); \quad (\mu, \Sigma) \to (A\mu + B, A\Sigma A^T)$$

with $A$ arbitrary $p \times p$ nonsingular matrix and $B$ arbitrary $p \times 1$ vector, the best affine equivariant estimator ($BAEE$) is

$$\hat{\Sigma}_1^{-1} = \frac{(n-p-4)(n-p-1)}{n-2} S^{-1}.$$

*This work is partially supported by the Hong Kong Polytechnic University Internal Research Grant No. G-S598 and by the NSF of Educational Committee of Jiangsu Province.

He also proved that the $BAEE$ $\hat{\Sigma}_1^{-1}$ is inadmissible and gave the following Haff-type improved estimator,

$$\hat{\Sigma}_2^{-1} = \frac{(n-p-4)(n-p-1)}{n-2}[S^{-1} + vt(v)I]$$

where $v = (trS)^{-1}$ and $t(v)$ is an absolutely continuous function such that $0 < t(v) \leq 2(p-1)/(n-p)$ and $t'(v) \leq 0$. Consider the following lower–triangular affine transformation group $\mathcal{C}$:

$$(1.1) \qquad (\bar{X}, S) \to (A\bar{X} + B, ASA^T); \quad (\mu, \Sigma) \to (A\mu + B, A\Sigma A^T)$$

where $A$ is a $p \times p$ lower–triangular matrix and $B$ is a $p \times 1$ vector. It can be easily shown that an equivariant estimator of $\Sigma^{-1}$ under the group $\mathcal{C}$ has the form $(K^T)^{-1}\Delta K^{-1}$ where $K$ is a lower–triangular matrix with positive diagonal elements such that $KK^T = S$ and $\Delta$ is an arbitrary diagonal matrix whose elements do not depend on $S$. Olkin and Selliah (1977) and Sharma and Krishnamoorthy (1983) have derived the best lower-triangular equivariant estimator ($BLEE$) of $\Sigma^{-1}$ for the case $p = 2$. For general $p$, the $BLEE$ of $\Sigma^{-1}$ is very complicated. Krishnamoorthy and Gupta (1989) gave the related expressions for determining the diagonal elements of $\Delta$. Because the lower–triangular affine transformation group $\mathcal{C}$ is the subgroup of the affine transformation group $\mathcal{A}$, the $BLEE$ of $\Sigma^{-1}$ is an improved estimator of the $BAEE$. In addition, since the lower–triangular affine transformation group $\mathcal{C}$ is solvable, from Kiefer (1957), the $BLEE$ of $\Sigma^{-1}$ is minimax. The admissibility of the $BLEE$ of $\Sigma^{-1}$ with respect to the squared loss $L_1$ is an interesting problem in statistical decision theory. Zhou and Sun (1999) proved that it is inadmissible for $p = 2$.

In this paper we consider the entropy loss

$$(1.2) \qquad L_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \text{tr}(\hat{\Sigma}^{-1}\Sigma) - \log|\hat{\Sigma}^{-1}\Sigma| - p.$$

Sinha and Ghosh (1987) showed that under the entropy loss $L_2$, the $BAEE$ of $\Sigma^{-1}$ is $\hat{\Sigma}_3^{-1} = (n-p-2)S^{-1}$ and they proved that the following Stein truncated type estimator

$$\hat{\Sigma}_4^{-1} = \begin{cases} (n-p-1)(S + n\bar{X}\bar{X}^T)^{-1}, & \text{if } \; n\bar{X}^T S^{-1}\bar{X} \leq (n-p-2)^{-1} \\ (n-p-2)S^{-1}, & \text{otherwise} \end{cases}$$

improves on the $BAEE$ $\hat{\Sigma}_3^{-1}$. Here we assume $n > p + 2$ throughout this paper. As pointed out by Krishnamoorthy and Gupta (1989), no Haff–type improved estimator of $\Sigma^{-1}$ is available. Sharma and Krishnamoorthy (1983) have derived the $BLEE$ of $\Sigma^{-1}$ for the case $p = 2$. For general $p$, the $BLEE$ of $\Sigma^{-1}$ is also derived by Krishnamoorthy and Gupta (1989), that is, $\hat{\Sigma}_5^{-1} = (K^T)^{-1}\Delta_2 K^{-1}$ where $K$ is as described above (after (1.1)) and $\Delta_2 = \text{diag}(\delta_{p1}, \ldots, \delta_{pp})$ with

$$(1.3) \qquad \delta_{pi} = \frac{(n-i-2)(n-i-1)}{(n-2)}, \quad i = 1, 2, \ldots, p.$$

Similarly, The $BLEE$ $\hat{\Sigma}_5^{-1}$ improves on the $BAEE$ $\hat{\Sigma}_3^{-1}$ and is also minimax.

The main purpose of this paper is to discuss the admissibility of the $BLEE$ $\hat{\Sigma}_5^{-1}$ with respect to the entropy loss $L_2$ and find an improved estimator. In the next section, an improved estimator of $BLEE$ $\hat{\Sigma}_5^{-1}$ is constructed from the class of equivariant estimators with respect to the lower–triangular scale transformation group $\mathcal{D}$ below, which proves the inadmissibility of the $BLEE$ $\hat{\Sigma}_5^{-1}$ with respect to the entropy loss $L_2$.

## 2. Main result

Let $\tilde{X} = \sqrt{n}\bar{X} = \sum_{i=1}^{n} X_i/\sqrt{n}$. Then $\tilde{X} \sim N_p(\sqrt{n}\mu, \Sigma)$ and $\tilde{X}$ is also independent of $S$. Let

$$VV^T = \sum_{i=1}^{n} X_i X_i^T \tag{2.1}$$

where $V$ is a lower-triangular matrix with positive diagonal elements. Then $VV^T = S + \tilde{X}\tilde{X}^T$ and $V^{-1}K$ is also lower-triangular with positive diagonal elements. Moreover, it is easy to show that $V^{-1}K \cdot (V^{-1}K)^T = I - (V^{-1}\tilde{X})(V^{-1}\tilde{X})^T$. Suppose $V^{-1}K = (y_{ij})_{p\times p}$ and $V^{-1}\tilde{X} = Y = (y_1, \ldots, y_p)^T$. Then we have

$$\begin{cases} y_{ij} = 0 \quad (i < j) \\ y_{11} = \sqrt{1 - y_1^2} \\ y_{i1} = -\dfrac{y_i y_1}{\sqrt{1-y_1^2}} \quad (i > 1) \\ y_{ii} = \dfrac{\sqrt{1 - \sum_{k=1}^{i} y_k^2}}{\sqrt{1 - \sum_{k=1}^{i-1} y_k^2}} \quad (i > 1) \\ y_{ij} = -\dfrac{y_i y_j}{\sqrt{1 - \sum_{k=1}^{j-1} y_k^2}\sqrt{1 - \sum_{k=1}^{j} y_k^2}} \quad (i > j > 1). \end{cases} \tag{2.2}$$

Let

$$(h_{ij})_{p\times p} = V^{-1}K \cdot (\Delta_2)^{-1} \cdot (V^{-1}K)^T = (y_{ij})_{p\times p} \operatorname{diag}\left(\frac{1}{\delta_{p1}}, \ldots, \frac{1}{\delta_{pp}}\right)(y_{ij})_{p\times p}^T.$$

Then from (2.2), we can easily obtain

$$\begin{cases} h_{11} = \dfrac{1 - y_1^2}{\delta_{p1}}, \\ h_{i1} = -\dfrac{y_i y_1}{\delta_{p1}}, \quad (i > 1) \\ h_{ii} = y_i^2 \left[\dfrac{y_1^2}{\delta_{p1}(1 - y_1^2)} + \cdots \right. \\ \qquad \left. + \dfrac{y_{i-1}^2}{\delta_{p,i-1}(1 - y_1^2 - \cdots - y_{i-2}^2)(1 - y_1^2 - \cdots - y_{i-1}^2)}\right] \\ \qquad + \dfrac{1 - y_1^2 - \cdots - y_i^2}{\delta_{pi}(1 - y_1^2 - \cdots - y_{i-1}^2)}, \quad (i > 1) \\ h_{ij} = y_i y_j \left[\dfrac{y_1^2}{\delta_{p1}(1 - y_1^2)} + \cdots \right. \\ \qquad + \dfrac{y_{j-1}^2}{\delta_{p,j-1}(1 - y_1^2 - \cdots - y_{j-2}^2)(1 - y_1^2 - \cdots - y_{j-1}^2)} \\ \qquad \left. - \dfrac{1}{\delta_{pj}(1 - y_1^2 - \cdots - y_{j-1}^2)}\right], \quad (i > j > 1) \end{cases} \tag{2.3}$$

and

$$(2.4)\qquad \begin{aligned}\hat{\Sigma}_5^{-1} &= (V^{-1})^T \cdot (K^{-1}V)^T \Delta_2 (K^{-1}V) \cdot V^{-1} = (V^{-1})^T \cdot \left(h_{ij}\right)_{p\times p}^{-1} \cdot V^{-1} \\ &\hat{=} (V^{-1})^T \cdot \left(d_{ij}\right)_{p\times p} \cdot V^{-1}\end{aligned}$$

where $\hat{=}$ stands for definition. From (2.3), it follows that

$$(2.5)\qquad (d_{ij})_{p\times p} = \Delta_2 \quad \text{if} \quad Y = (0,\dots,0)^T.$$

Let $(w_{ij})_{p\times p} = (y_{ij})_{p\times p}^{-1}$, then $(d_{ij})_{p\times p} = (w_{ij})_{p\times p}^T \cdot \mathrm{diag}(\delta_{p1},\dots,\delta_{pp}) \cdot (w_{ij})_{p\times p}$. Similar to $(y_{ij})_{p\times p}$, $(w_{ij})_{p\times p}$ is also lower triangular. Obviously,

$$\begin{cases} w_{ij} = y_{ij}^{-1} \\ (w_{uv})_{1\le u,v\le i-1} = (y_{uv})_{1\le u,v\le i-1}^{-1} \\ (w_{i1},\dots,w_{i,i-1}) = -w_{ii}\cdot (y_{i1},\dots,y_{i,i-1}) \cdot (w_{uv})_{1\le u,v\le i-1}. \end{cases}$$

From (2.2), it follows that

$$w_{ii} = \frac{\sqrt{1-\sum_{k=1}^{i-1} y_k^2}}{\sqrt{1-\sum_{k=1}^{i} y_k^2}}.$$

Using inductive method, it can be shown that $w_{ij}$ has the form of $y_i y_j g(y_1^2,\dots,y_i^2)$. Thus $d_{ij}$'s have the following form: $d_{ii} = g_{ii}(y_1^2,\dots,y_p^2)$ for $i = 1,\dots,p$ and $d_{ij} = y_i y_j g_{ij}(y_1^2,\dots,y_p^2)$, with $g_{ij} = g_{ji}$, for $i,j = 1,\dots,p; i \neq j$.

Consider the lower–triangular scale transformation group $\mathcal{D}$:

$$(\bar{X}, S) \to (A\bar{X}, ASA^T); \quad (\mu,\Sigma) \to (A\mu, A\Sigma A^T)$$

where $A$ is a $p\times p$ lower—triangular matrix, which is obviously a subgroup of the lower–triangular affine transformation group $\mathcal{C}$ defined in (1.1). Similarly to Wang (1984) for estimating the covariance matrix $\Sigma$, we consider the estimator of the form $\hat{\Sigma}^{-1} = (V^{-1})^T\cdot D(Y)\cdot V^{-1}$ where $D(Y)$ is an arbitrary $p\times p$ positive definite matrix with diagonal elements $d_{ii}(Y)$ of the form $f_{ii}(y_1^2,\dots,y_p^2)$, $i = 1,\dots,p$, and off–diagonal elements $d_{ij}(Y)$ of the form $y_i y_j f_{ij}(y_1^2,\dots,y_p^2)$, $i \neq j$. It can be easily shown that the estimator of this form is an equivariant estimator of $\Sigma^{-1}$ under the group $\mathcal{D}$. Obviously, the class $C_2$ of equivariant estimators of $\Sigma^{-1}$ under the lower–triangular scale transformation group $\mathcal{D}$ is larger than the class $C_1$ of equivariant estimators of $\Sigma^{-1}$ under the lower—triangular affine transformation group $\mathcal{C}$ because $\mathcal{D}$ is a subgroup of $\mathcal{C}$. So the *BLEE* $\hat{\Sigma}_5^{-1}$ is in $C_2$. In the following we will find an improved estimator of the *BLEE* $\hat{\Sigma}_5^{-1}$ in the class $C_2$.

Let $\Sigma = GG^T$, where $G$ is a lower-triangular matrix with positive diagonal elements. Since we only consider the estimators within the class $C_2$, we can assume $S \sim W_p(n-1, I_p)$ and $\tilde{X} \sim N_p(\beta, I_p)$ without loss of generality, where $\beta = (\beta_1,\dots,\beta_p)^T = \sqrt{n}G^{-1}\mu$. The joint probability density function of $(S,\tilde{X})$ is

$$p(S,\tilde{X}) = \begin{cases} \dfrac{|S|^{(n-p-2)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(S)\right\}}{2^{(n-1)p/2}\pi^{p(p-1)/4}\prod_{i=1}^p \Gamma\left(\frac{n-i}{2}\right)} \cdot \dfrac{\exp\left\{-\frac{1}{2}(\tilde{X}-\beta)^T(\tilde{X}-\beta)\right\}}{(2\pi)^{p/2}}, & \text{if } S \text{ is positive definite;} \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to show that the joint probability density function of $(V, Y)$ is

$$p(V,Y) = \begin{cases} c \cdot \prod_{i=1}^{p} v_{ii}^{n-i} \exp\left\{-\frac{1}{2}\sum_{i\geq j} v_{ij}^2\right\} \cdot \exp\{\beta^T VY\}(1-Y^TY)^{(n-p-2)/2} \\ \qquad\qquad \text{if } \; Y^TY < 1 \text{ and } V_{ii} > 0 \quad (1 \leq i \leq p); \\ 0, \qquad \text{otherwise} \end{cases} \tag{2.6}$$

where $c = [2^{(n-2)p/2}\pi^{p(p+1)/4}\prod_{i=1}^{p}\Gamma(\frac{n-i}{2})\exp\{-\frac{1}{2}\beta^T\beta\}]^{-1}$.

From (2.6) we can see that given $Y$, all the $v_{ij}$'s are independent.

Now we define the following new estimator of $\Sigma^{-1}$:

$$\hat{\Sigma}_6^{-1} = (V^{-1})^T \cdot D_1(Y) \cdot V^{-1} \tag{2.7}$$

where $V$ is defined by (2.1) and $D_1(Y)$ is obtained from $(d_{ij})_{p\times p}$ in (2.4) by only substituting its diagonal component $d_{11}$ with $d$, a function of $y_1^2, \ldots, y_p^2$. We will choose a suitable value for $d$ such that $D_1(Y)$ is positive definite and $\hat{\Sigma}_6^{-1}$ is better than the *BLEE* $\hat{\Sigma}_5^{-1}$ with respect to the entropy loss $L_2$. Note that if $d$ is a function of $y_1^2, \ldots, y_p^2$ and $D_1(Y)$ is positive definite, then $\hat{\Sigma}_6^{-1}$ is in the class $C_2$ of equivariant estimators of $\Sigma^{-1}$ under the lower–triangular scale transformation group $\mathcal{D}$.

The loss of $\hat{\Sigma}_6^{-1} = (V^{-1})^T \cdot D_1(Y) \cdot V^{-1}$ is

$$\begin{aligned} L_2(d) &\triangleq L_2(\hat{\Sigma}_6^{-1}, I_p) = \operatorname{tr}((V^{-1})^T D_1(Y) V^{-1}) - \log|(V^{-1})^T D_1(Y) V^{-1}| - p \\ &= \operatorname{tr}[(V^{-1})^T(D_1(Y) - dE_{11})V^{-1}] + d\cdot\operatorname{tr}((V^{-1})^T E_{11} V^{-1}) \\ &\quad - \log|D_1(Y)| + \log|VV^T| - p \\ &= d\cdot\frac{1}{v_{11}^2} - \log|D_1(Y)| + \operatorname{tr}[(V^{-1})^T(D_1(Y) - dE_{11})V^{-1}] + \log|VV^T| - p \end{aligned}$$

where $E_{11} = (e_{ij})_{p\times p}$ denotes the $p \times p$ matrix with $e_{11} = 1$ and all other elements zero. Hence the risk of $\hat{\Sigma}_6^{-1}$ is

$$R(d) \triangleq R(\hat{\Sigma}_6^{-1}, \theta) = \int L_2(d)p(V,Y)dVdY \triangleq \int_{Y^TY<1} R^Y(d)dY$$

where

$$R^Y(d) = \int L_2(d)p(V,Y)dV.$$

Since $p(V, Y)$ in (2.6) can also be written as

$$p(V,Y) = \begin{cases} c \cdot \prod_{i=1}^{p} v_{ii}^{n-i} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p} v_{ii}^2\right\} \prod_{i=1}^{p}\sum_{k=0}^{\infty}\frac{(\beta_i v_{ii} y_i)^k}{k!} \exp\left\{\frac{1}{2}\sum_{i>j}\beta_i^2 y_j^2\right\} \\ \qquad \times \exp\left\{-\frac{1}{2}\sum_{i>j}(v_{ij} - \beta_i y_j)^2\right\}(1-Y^TY)^{(n-p-2)/2} \\ \qquad\qquad \text{if } \; Y^TY < 1 \text{ and } v_{ii} > 0 \quad (1 \leq i \leq p); \\ 0, \qquad \text{otherwise,} \end{cases}$$

and $|D_1(Y)|$ is an even function of $y_1$, we have

$$R^Y(d) = B[d\cdot a(Y) - b(Y)\log|D_1(Y)|] + c(Y) \tag{2.8}$$

where $B$ is a positive-valued function of $\beta_1^2, \ldots, \beta_p^2$ and $y_1^2, \ldots, y_p^2$, i.e.,

$$\begin{aligned}
B &\hat{=} B(\beta_1^2, \ldots, \beta_p^2; y_1^2, \ldots, y_p^2) > 0, \\
a(Y) &= \sum_{k=0}^{\infty} \int_{v_{pp}>0} v_{11}^{-2} v_{11}^{n-1} \frac{(\beta_1 v_{11} y_1)^{2k}}{(2k)!} \exp\left\{-\frac{1}{2} v_{11}^2\right\} dv_{11} \\
&= \sum_{k=0}^{\infty} \frac{(\beta_1 y_1)^{2k}}{(2k)!} \int_0^{\infty} v_{11}^{n+2k-3} \exp\left\{-\frac{1}{2} v_{11}^2\right\} dv_{11} \\
&= \sum_{k=0}^{\infty} \frac{(\beta_1 y_1)^{2k}}{(2k)!} \int_0^{\infty} t^{(n+2k-2)/2-1} \exp\left\{-\frac{t}{2}\right\} \cdot \frac{1}{2} dt \\
&= \sum_{k=0}^{\infty} \frac{(\beta_1 y_1)^{2k}}{(2k)!} \Gamma\left(\frac{n+2k-2}{2}\right) 2^{(n+2k-4)/2} > 0, \\
b(Y) &= \sum_{k=0}^{\infty} \int_{v_{11}>0} v_{11}^{n-1} \frac{(\beta_1 v_{11} y_1)^{2k}}{(2k)!} \exp\left\{-\frac{1}{2} v_{11}^2\right\} dv_{pp} \\
&= \sum_{k=0}^{\infty} \frac{(\beta_1 y_1)^{2k}}{(2k)!} \Gamma\left(\frac{n+2k}{2}\right) 2^{(n+2k-2)/2} > 0
\end{aligned}$$

and $c(Y)$ is independent of $d$. Here note that all odd power terms of $y_1$ vanish because the next domain of integration, $\{(y_1, \ldots, y_p) : Y^T Y < 1\}$, is symmetric about the origin.

Let $\Omega$ be a subset of $R^p$ such that each element $Y$ satisfies the following conditions

$$\text{(2.9)} \qquad \begin{cases} d_{11} < n-2 \\ Y'Y < 1 \end{cases}$$

and define

$$\text{(2.10)} \qquad d_0 = \begin{cases} n-2, & \text{if } Y \in \Omega, \\ d_{11}, & \text{otherwise.} \end{cases}$$

Then we have the following main result:

THEOREM. *Under the entropy loss $L_2$ in (1.2), the BLEE of the multivariate normal precision matrix $\Sigma^{-1}$, $\hat{\Sigma}_5^{-1} = (K^{-1})^T \Delta_2 K^{-1}$, is inadmissible, where $K$ is lower-triangular with positive diagonal elements such that $KK^T = S$ and $\Delta_2 =$ diag$(\delta_{p1}, \ldots, \delta_{pp})$ with $\delta_{pi}$'s as in (1.3). Furthermore, $\hat{\Sigma}_6^{-1} = (V^{-1})^T D_1(Y) V^{-1}$ improves on BLEE $\hat{\Sigma}_5^{-1}$, where $D_1(Y)$ is seen in (2.7) with $d = d_0$ given by (2.10) and $V$ is defined by (2.1).*

PROOF. It only needs to show $R(\hat{\Sigma}_6^{-1}, \beta) \leq R(\hat{\Sigma}_5^{-1}, \beta)$ or $R(d_0) \leq R(d_{11})$ for all $\beta$ and with strict inequality for some $\beta$. Thus it suffices to prove

$$\text{(2.11)} \qquad R^Y(d_0) \leq R^Y(d_{11})$$

for all $\beta \in R^p$, $Y \in \{Y : Y^T Y < 1\}$, and with strict inequality for some $\beta$ and a subset of $\{Y : Y^T Y < 1\}$ with positive Lebesgue measure.

Let $\alpha = (d_{21}, \ldots, d_{p-1,1})^T$. From (2.8), the derivative of $R^Y(d)$ with respect to $d$ is

$$\text{(2.12)} \qquad [R^Y(d)]' = B\left[a(Y) - b(Y) \cdot \frac{|D_1(Y)|'}{|D_1(Y)|}\right]$$

and the derivative of $|D_1(Y)|$ with respect to $d$, $|D_1(Y)|' = |D_1^*(Y)|$ is independent of $d$, where $D_1^*(Y)$ is the $(p-1) \times (p-1)$ matrix obtained from $D_1(Y)$ by eliminating the first column and the first row. Thus,

$$|D_1(Y)| = \begin{vmatrix} d & \alpha^T \\ \alpha & D_1^*(Y) \end{vmatrix} = \begin{vmatrix} d - \alpha^T[D_1^*(Y)]^{-1}\alpha & 0 \\ \alpha & D_1^*(Y) \end{vmatrix}$$
$$= |D_1^*(Y)| \cdot (d - \alpha^T[D_1^*(Y)]^{-1}\alpha).$$

Hence, from (2.12), it follows that

$$[R^Y(d)]' = B\left[a(Y) - \frac{b(Y)}{d - \alpha^T[D_1^*(Y)]^{-1}\alpha}\right]$$
$$= B\sum_{k=0}^{\infty} \frac{(\beta_1 y_1)^{2k}}{(2k)!}\Gamma\left(\frac{n+2k-2}{2}\right)2^{(n+2k-4)/2}\left[1 - \frac{n+2k-2}{d - \alpha^T[D_1^*(Y)]^{-1}\alpha}\right].$$

Therefore, a sufficient condition for $[R^Y(d)]' \le 0$ is $0 < d - \alpha^T[D_1^*(Y)]^{-1}\alpha \le n-2$.

For any $Y \notin \Omega$, from the definition of $d_0$, we have $d_0 = d_{11}$ , and hence $R^Y(d_0) = R^Y(d_{11})$. For any $Y \in \Omega$, $d_0 = n-2$. It is obvious that $d_{11} - \alpha^T[D_1^*(Y)]^{-1}\alpha > 0$ because the matrix $(d_{ij})_{p \times p}$ is positive definite. Thus it follows that $[R^Y(d)]' \le 0$ for $d \in [d_{11}, n-2]$, hence $R^Y(d_0) = R^Y(n-2) \le R^Y(d_{11})$. Also, it is easy to show that $D_1(Y)$ is positive definite when $d = d_0$ as in (2.10). In addition, from (2.5) and (2.9), $Y = (0,\ldots,0)^T$ is an inner point of $\Omega$ because $d_{11} = \delta_{p1} = n-3$ when $Y = (0,\ldots,0)^T$, and this can assure that the Lebesgue measure of $\Omega$ is greater than zero; thus we can easily show that (2.11) holds strictly for some $\beta \in R^p$ and a subset of $\{Y : Y'Y < 1\}$ with positive Lebesgue measure, which completes the proof. □

Since $\hat{\Sigma}_5^{-1}$ is minimax, our improved estimator $\hat{\Sigma}_6^{-1}$ is also minimax.

To confirm the theoretical results in this paper, we carried out some Monte Carlo studies and present some of the simulation results below. First we note that because $\hat{\Sigma}_5^{-1}$ is *BLEE*, its risk is constant. For $\hat{\Sigma}_6^{-1}$, as mentioned above, it is in the class $C_2$ of equivariant estimators of $\Sigma^{-1}$ under the lower—triangular scale transformation group $\mathcal{D}$ with respect to the entropy loss $L_2$. Hence without loss of generality we can assume $X_1,\ldots,X_n$ to be a random sample from a multivariate normal distribution $N_p(\mu, I_p)$ where $\mu = (\mu_1,\ldots,\mu_p)^T \in R^p$. In the following tables, simulated entropy risks of $\hat{\Sigma}_5^{-1}$ and $\hat{\Sigma}_6^{-1}$ are provided for $p = 2$, $n = 5, 10$, and selected values of $\mu = (\mu_1, \mu_2)^T$.

Table 1. Simulated risks for $p = 2$, $n = 5$.

Risk of $\hat{\Sigma}_5^{-1}$: 1.6131
Risk of $\hat{\Sigma}_6^{-1}$:

| | $\mu_2$ | | | | |
|---|---|---|---|---|---|
| $\mu_1$ | $-1$ | $-0.5$ | 0 | 0.5 | 1 |
| $-1$ | 1.6039 | 1.6017 | 1.6005 | 1.6013 | 1.6036 |
| $-0.5$ | 1.5860 | 1.5822 | 1.5807 | 1.5823 | 1.5859 |
| 0 | 1.5801 | 1.5776 | 1.5775 | 1.5784 | 1.5805 |
| 0.5 | 1.5855 | 1.5819 | 1.5805 | 1.5818 | 1.5851 |
| 1 | 1.6016 | 1.5988 | 1.5973 | 1.5986 | 1.6013 |

Table 2. Simulated risks for $p = 2$, $n = 10$.

Risk of $\hat{\Sigma}_5^{-1}$: 0.4254
Risk of $\hat{\Sigma}_6^{-1}$:

| $\mu_1$ \ $\mu_2$ | $-1$ | $-0.5$ | 0 | 0.5 | 1 |
|---|---|---|---|---|---|
| $-1$ | 0.4253 | 0.4252 | 0.4251 | 0.4252 | 0.4253 |
| $-0.5$ | 0.4239 | 0.4234 | 0.4231 | 0.4234 | 0.4239 |
| 0 | 0.4214 | 0.4212 | 0.4211 | 0.4212 | 0.4216 |
| 0.5 | 0.4241 | 0.4236 | 0.4233 | 0.4236 | 0.4242 |
| 1 | 0.4253 | 0.4252 | 0.4252 | 0.4252 | 0.4253 |

From Tables 1–2, we can see that $\hat{\Sigma}_6^{-1}$ provides substantial improvement over $\hat{\Sigma}_5^{-1}$ in terms of the risk under the entropy loss for $\mu$ close to the origin, but the improvement diminishes as $\mu$ moves away from the origin.

*Remark.* The estimator $\hat{\Sigma}_5^{-1}$ can also be improved based on the idea of averaging existing estimators introduced by Stein (1956). Takemura (1984) extends this idea to averaging over orthogonal transformations and establishes an orthogonally invariant estimator which improves the *BLEE* of the covariance matrix $\Sigma$. This approach is applicable to $\hat{\Sigma}_5^{-1}$ as well. We do not elaborate much further on this approach here as it is well known to the researchers of this field.

## Acknowledgements

We are grateful to the editor and two anonymous referees for their valuable comments and suggestions which contribute to the improvment of this paper.

## References

Haff, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity, *Ann. Statist.*, **7**, 1264–1276.

Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes, *Ann. Math. Statist.*, **28**, 573–601.

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

Kubokawa, T. (1989). Improved estimation of a covariance matrix under quadratic loss, *Statist. Probab. Lett.*, **8**, 69–71.

Olkin, I. and Selliah, J. (1977). Estimating covariance in a multivariate normal distribution, *Statistical Decision Theory and Related Topics, II* (ed. S. S. Gupta), 313–326, Academic Press, New York.

Pal, N. (1988). Decision-theoretic estimation of generalized variance and generalized precision, *Comm. Statist. Theory Methods*, **17**, 4221–4230.

Sharma, D. and Krishnamoorthy, K. (1983). Orthogonal equivariant estimator bivariate normal covariance matrix and precision matrix, *Calcutta Statist. Assoc. Bull.*, **32**, 23–45.

Sinha, B. K. and Ghosh, M. (1987). Inadmissibility of the best equivariant estimators of the variance–covariance matrix, the precision matrix and the generalized variance under entropy loss, *Statist. Decisions*, **5**, 201–227.

Stein, C. (1956). Some problems in multivariate analysis, Part 1, *Tech. Report*, No. 6, Department of Statistics, Stanford University, California.

Sugiura, N. (1988). A class of improved estimators of powers of the generalized variance and precision under squared loss, *Statistical Theory and Data Analysis II* (ed. K. Matusita), 421–428, Elsevier, North–Holland, Amsterdam.

Sun, X. Q. (1998). Improved estimation of the generalized precision under the squared loss, *Comm. Statist. Theory Methods*, **23**, 2725–2742.

Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance mtrix of a multivariate normal population, *Tsukuba J. Math.*, **8**, 367–376.

Wang, J. L. (1984), Improvement on the best affine invariant estimation of the covariance matrix, *Acta Math. Appl. Sinica*, **7**, 219–234 (in Chinese).

Zhou, X. and Sun, X. Q. (1999). Estimation of the bivariate normal precision matrix under the squared loss, Tech. Report, No. 15, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong.