ISSN 1066-5307, Mathematical Methods of Statistics, 2024, Vol. 33, No. 1, pp. 12–25. © Allerton Press, Inc., 2024.

Truncated Estimators for a Precision Matrix

Anis M. Haddouche[1*] and Dominique Fourdrinier[2]**

[1]IMT Atlantique, LABSTICC, 655 Avenue du Technopôle, Plouzané, 29280 France

[2]Université de Normandie, Normandie Univ, Unirouen, Unihavre, Insa Rouen, LITIS, Avenue de l'Université, BP 12, Saint-Étienne-du-Rouvray, 76801 France

Received November 11, 2022; revised September 21, 2023; accepted October, 2023

Abstract—In this paper, we estimate the precision matrix Σ^{-1} of a Gaussian multivariate linear regression model through its canonical form $(Z^T, U^T)^T$ where Z and U are respectively an $m \times p$ and an $n \times p$ matrices. This problem is addressed under the data-based loss function tr $[(\hat{\Sigma}^{-1} - \Sigma^{-1})S]^2$, where $\hat{\Sigma}^{-1}$ estimates Σ^{-1}, for any ordering of m, n and p, in a unified approach. We derive estimators which, besides the information contained in the sample covariance matrix $S = U^T U$, use the information contained in the sample mean Z. We provide conditions for which these estimators improve over the usual estimators aS^+ where a is a positive constant and S^+ is the Moore-Penrose inverse of S. Thanks to the role of Z, such estimators are also improved by their truncated version.

DOI: 10.3103/S1066530724700029

Keywords: Data-based loss; precision matrix; statistical decision theory; truncated estimators; high—dimensional statistics

1. INTRODUCTION

Consider the canonical form $(Z^T, U^T)^T$ of the multivariate linear regression model, where Z and U are respectively an $m \times p$ and an $n \times p$ mutually independent random matrices distributed as

$$Z \sim \mathcal{N}_{m\times p}(\theta, I_m \otimes \Sigma) \quad \text{and} \quad U \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$$

so that $(Z^T, U^T)^T$ has density

$$(2\pi)^{-(m+n)p/2}|\Sigma|^{-(m+n)/2} \exp[-\text{tr}\,\{(z-\theta)\Sigma^{-1}(z-\theta)^T + u\Sigma^{-1}u^T\}/2]. \tag{1}$$

Here θ is the unknown mean matrix and Σ is the unknown positive definite covariance matrix.

It is worth pointing out that, in the case where $p \leq n$, the $p \times p$ sample covariance matrix $S = U^T U$ is invertible and has a Wishart distribution. Srivastava [13] gave a generalization of the Wishart distribution, called the singular Wishart, to the case where $p > n$, that is, when S is non-invertible. In the following, for both singular and non-singular cases, we use the notation $S \sim \mathcal{W}_p(n, \Sigma)$ and S^+ holds for the Moore—Penrose inverse of S, which coincides with the regular inverse S^{-1} in the case where $p \leq n$.

The model in (1) has been used by various authors in the literature, e.g., Tsukuma and Kubokawa [17] for the estimation of the scale matrix Σ, and Tsukuma and Kubokawa [16] for the estimation of the mean matrix θ. We refer to Tsukuma and Kubokawa [18] for more details about the development of this canonical form. Note that an extension to the case where the density is elliptically symmetric has been considered by Haddouche et al. [7] and by Canu and Fourdrinier [2].

In this paper, we deal with the problem of estimating the precision matrix Σ^{-1}, based on the sufficient statistic (Z, S), where the performance of any estimator $\hat{\Sigma}^{-1}$ is assessed under the loss function

$$L\Big(\Sigma^{-1}, \hat{\Sigma}^{-1}\Big) = \text{tr}\left[\Big(\hat{\Sigma}^{-1} - \Sigma^{-1}\Big)S\right]^2 \tag{2}$$

*E-mail: mohamed.anis.haddouche@gmail.com

**E-mail: Dominique.Fourdrinier@univ-rouen.fr

and its associated risk

$$R\Big(\Sigma^{-1}, \hat{\Sigma}^{-1}\Big) = E_{\theta,\Sigma}\left[\text{tr}\left[\Big(\hat{\Sigma}^{-1} - \Sigma^{-1}\Big) S\right]^2\right], \tag{3}$$

where $E_{\theta,\Sigma}$ denotes the expectation with respect to the density (1). The loss in (2) differs from the usual quadratic loss tr $[(\hat{\Sigma}^{-1} - \Sigma^{-1})]^2$ in so far as, beside the dependence of $\hat{\Sigma}^{-1}$ with respect to S, the matrix S intervenes in the loss structure. Following Fourdrinier and Strawderman [3] we call such a loss a data-based loss. For various estimation problems, analogous data-based losses have been used by many authors such as Fourdrinier et al. [5], Kubokawa and Srivastava [10], and Boukehil et al. [1] since these losses turned out to allow evaluation of more orthogonally invariant estimators (see Takemura [15]) than the usual quadratic loss.

For the extended class of estimators that we consider, we will see that our loss in (2) is more suitable (see Remarks 1 and 2 below for more details).

It is shown in Section A.3 of Appendix A that the optimal estimator, which minimizes the risk (3), among the class of the usual estimators of the form aS^+ where a is a positive constant, is

$$\hat{\Sigma}_o^{-1} = a_o S^+ \quad \text{with} \quad a_o = n \vee p, \tag{4}$$

where $n \vee p$ is the maximum between n and p. Note that, the estimator $\hat{\Sigma}_o^{-1}$ in (4) depends only on the statistic S.

There exist estimators better than $\hat{\Sigma}_o^{-1}$ depending only on S, such as the orthogonally invariant estimators considered by Kubokawa and Srivastava [10] and Fourdrinier et al. [4] (see Remark A.4 in the Appendix A). We enlarge the class of such improved estimators with estimators that also involve the information contained in the statistic Z. This dependence on Z allows also the construction of new estimators, called truncated estimators, which improve over the "non-truncated" estimators. This estimation problem is addressed for any ordering of m, n, and p, in a unified approach.

These estimators parallel those considered by Kubokawa and Srivastava [9], Kubokawa and Tsai [11], and Tsukuma and Kubokawa [17] in the context of estimating the covariance matrix Σ. While the original idea of using Z is due to Stein [14], to our knowledge only Sinha and Ghosh [12] considered such estimators for estimating the precision matrix under the Stein loss function

$$L\big(\Sigma^{-1}, \hat{\Sigma}^{-1}\big) = \text{tr}\,(\Sigma\hat{\Sigma}^{-1}) - \log|\Sigma\hat{\Sigma}^{-1}| - p$$

and in the case where S is invertible.

This paper is structured as follows. In Section 2, dominance results are provided in a unified approach with respect to any ordering of m, n, and p, under the risk function (3). We first show that $\hat{\Sigma}_o^{-1}$ is dominated by a class of estimators involving the statistic Z. Secondly, we give a truncation rule which leads to derive estimators that improve over the latter. Examples of such estimators illustrate the results. In Section 3, we complete the investigation of the performance of these estimators through a numerical study. In Section 4, we give concluding remarks and some perspectives. Finally, an appendix contains technical results which have been used in the development of the paper.

2. MAIN RESULTS

2.1. Preliminaries

Consider the model in (1). Let

$$S = HLH^T \tag{5}$$

be the eigenvalue decomposition of the sample covariance matrix $S = U^TU$ where the matrix $L = \text{diag}(l_1 > ... > l_i > ... > l_{n\wedge p} > 0)$ is the $(n \wedge p) \times (n \wedge p)$ diagonal matrix of the positive eigenvalues of S and H the $p \times (n \wedge p)$ semi-orthogonal matrix ($H^TH = I_{n\wedge p}$) of the corresponding eigenvectors (see Srivastava [13] for more details). In order to construct a new class of estimators, we combine the information on the unknown scatter matrix Σ and the mean matrix θ, contained respectively in the matrices S and Z. To this end, through Lemma A.3 and its proof, we rely on the thin singular value decomposition of $ZHL^{-1/2}$ defined as

$$ZHL^{-1/2} = R\Lambda^{1/2}O^T,$$

where

$$\Lambda = \text{diag}(\lambda_1 > \cdots > \lambda_i > \cdots > \lambda_k > 0) \tag{6}$$

is a $k \times k$ diagonal matrix with $k = n \wedge m \wedge p$, O is an $(n \wedge p) \times k$ semi-orthogonal matrix ($O^T O = I_k$) and R is also an $m \times k$ semi-orthogonal matrix ($R^T R = I_k$). See Subsection 2.5.4 of Golub and van Loan [6]. In particular, we deal with the simultaneous diagonalization of S in (5) and

$$W = Z^T Z. \tag{7}$$

2.2. An Extended Class of Estimators

The estimators considered in this section are based on the information contained in both the statistics S and Z. They are of the form

$$\hat{\Sigma}_{\Psi}^{-1} = a_o \left(S^+ + (Q^-)^T \Psi(\Lambda) Q^- \right), \tag{8}$$

where $S^+ = HL^{-1}H^T$ is the Moore–Penrose inverse of S, $\Psi(\Lambda)$ is a $k \times k$ diagonal matrix such that its elements are absolutely continuous functions of the λ_i's in (6) and

$$Q^- = O^T L^{-1/2} H^T,$$

which is, according to Lemma A.3, a reflexive generalized inverse of

$$Q = HL^{1/2}O.$$

We emphasize that, when $k = n \wedge p$, the Moore–Penrose inverse S^+ of S equals $(Q^-)^T Q^-$ since O is orthogonal. In that case, the class of estimators in (8) can be re-written as

$$\hat{\Sigma}_{\Psi}^{-1} = a_o (Q^-)^T \left(I_{n\wedge p} + \Psi(\Lambda) \right) Q^-,$$

which parallels the class considered by Kubokawa and Tsai [11] for estimating Σ.

The following theorem gives an improvement and a dominance result, with respect to the risk in (3), of the alternative estimators $\hat{\Sigma}_{\Psi}^{-1}$ in (8) over the usual estimator $\hat{\Sigma}_{o}^{-1}$ in (4). The improvement result consists in showing that the risk difference between $\hat{\Sigma}_{\Psi}^{-1}$ and $\hat{\Sigma}_{o}^{-1}$ satisfies

$$\Delta_{\theta,\Sigma}(\Psi) = R\left(\hat{\Sigma}_{\Psi}^{-1}, \Sigma^{-1}\right) - R\left(\hat{\Sigma}_{o}^{-1}, \Sigma^{-1}\right) \leq 0 \tag{9}$$

for any θ and any Σ. As for the dominance result, one has to show that, in addition to the fact that (9) is satisfied, this inequality is strict for some (θ, Σ).

Theorem 1. *Let $\Psi(\Lambda) = \text{diag}\left(\psi_1(\Lambda), \ldots, \psi_i(\Lambda), \ldots, \psi_k(\Lambda)\right)$ be a $k \times k$ diagonal matrix function of Λ in (6) such that, for any $i = 1, \ldots, k$, the function $\lambda_i \mapsto \psi_i\left(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_k\right)$ is absolutely continuous.*

The risk difference between $\hat{\Sigma}_{\Psi}^{-1}$ in (8) and $\hat{\Sigma}_{o}^{-1}$ in (4) is expressed as

$$\Delta_{\theta,\Sigma}(\Psi) = a_o E_{\theta,\Sigma}\big[\delta(\Psi)\big],$$

where $\delta(\Psi)$ equals

$$\sum_{i=1}^{k} \left\{ (n \vee p)\psi_i^2 - 2(2k - n \wedge p - 1)\psi_i + 4\lambda_i \frac{\partial \psi_i}{\partial \lambda_i} + 4 \sum_{j>i} \frac{\lambda_i \psi_i - \lambda_j \psi_j}{\lambda_i - \lambda_j} \right\}, \tag{10}$$

so that the random variable $a_o\delta(\Psi)$ appears as an unbiased estimator of this risk difference. Then $\hat{\Sigma}_{\Psi}^{-1}$ improves over (respectively dominates) $\hat{\Sigma}_{o}^{-1}$ as soon as $\delta(\Psi)$ is non positive (respectively negative).

Proof. Note that $\hat{\Sigma}_\Psi^{-1}$ can be rewritten as $\hat{\Sigma}_\Psi^{-1} = \hat{\Sigma}_o^{-1} + a_o (Q^-)^T \Psi(\Lambda) Q^-$ so that the risk at Σ^{-1} of $\hat{\Sigma}_\Psi^{-1}$ in (8) is

$$
\begin{aligned}
R\left(\Sigma^{-1}, \hat{\Sigma}_\Psi^{-1}\right) &= E_{\theta,\Sigma}\left[\text{tr}\left[\left(\hat{\Sigma}_o^{-1} - \Sigma^{-1} + a_o (Q^-)^T \Psi(\Lambda) Q^-\right) S\right]^2\right] \\
&= E_{\theta,\Sigma}\left[\text{tr}\left[\left(\hat{\Sigma}_o^{-1} - \Sigma^{-1}\right) S\right]^2\right] + a_o^2 E_{\theta,\Sigma}\left[\text{tr}\left((Q^-)^T \Psi(\Lambda) Q^- S\right)^2\right] \\
&\quad + 2a_o E_{\theta,\Sigma}\left[\text{tr}\left(\hat{\Sigma}_o^{-1} S (Q^-)^T \Psi(\Lambda) Q^- S\right)\right] \\
&\quad - 2a_o E_{\theta,\Sigma}\left[\text{tr}\left(\Sigma^{-1} S (Q^-)^T \Psi(\Lambda) Q^- S\right)\right],
\end{aligned}
$$

where

$$
E_{\theta,\Sigma}\left[\text{tr}\left[\left(\hat{\Sigma}_o^{-1} - \Sigma^{-1}\right) S\right]^2\right] = R\left(\Sigma^{-1}, \hat{\Sigma}_o^{-1}\right).
$$

Thus the risk difference between $\hat{\Sigma}_\Psi^{-1}$ and the optimal estimator $\hat{\Sigma}_o^{-1}$ equals

$$
\begin{aligned}
\Delta_{\theta,\Sigma}(\Psi) &= R\left(\Sigma^{-1}, \hat{\Sigma}_\Psi^{-1}\right) - R\left(\Sigma^{-1}, \hat{\Sigma}_o^{-1}\right) \\
&= a_o^2 E_{\theta,\Sigma}\left[\text{tr}\left((Q^-)^T \Psi(\Lambda) Q^- S\right)^2\right] + 2a_o E_{\theta,\Sigma}\left[\text{tr}\left(\hat{\Sigma}_o^{-1} S (Q^-)^T \Psi(\Lambda) Q^- S\right)\right] \\
&\quad - 2a_o E_{\theta,\Sigma}\left[\text{tr}\left(\Sigma^{-1} S (Q^-)^T \Psi(\Lambda) Q^- S\right)\right].
\end{aligned}
$$

Now we have

$$
\begin{aligned}
\text{tr}\left(\left\{(Q^-)^T \Psi(\Lambda) Q^- S\right\}^2\right) &= \text{tr}\left(\left\{S (Q^-)^T \Psi(\Lambda) Q^-\right\}^2\right) \\
&= \text{tr}\left(\left\{Q \Psi(\Lambda) Q^-\right\}^2\right) \quad \text{according to (A.10)} \\
&= \text{tr}\left(\left\{Q^- Q \Psi(\Lambda)\right\}^2\right) \\
&= \text{tr}\left(\left\{\Psi(\Lambda)\right\}^2\right) \quad \text{according to (A.9)},
\end{aligned}
$$

$$
\begin{aligned}
\text{tr}\left(\hat{\Sigma}_o^{-1} S (Q^-)^T \Psi(\Lambda) Q^- S\right) &= a_o \text{tr}\left(S^+ S (Q^-)^T \Psi(\Lambda) Q^- S\right) \\
&= a_o \text{tr}\left(S^+ Q \Psi(\Lambda) Q^- S\right) \quad \text{according to (A.10)} \\
&= a_o \text{tr}\left(Q^- S S^+ Q \Psi(\Lambda)\right) \\
&= a_o \text{tr}\left(Q^T S^+ Q \Psi(\Lambda)\right) \quad \text{according to (A.10)} \\
&= a_o \text{tr}\left(\Psi(\Lambda)\right) \quad \text{according to (A.12)}
\end{aligned}
$$

and

$$
\text{tr}\left(\Sigma^{-1} S (Q^-)^T \Psi(\Lambda) Q^- S\right) = \text{tr}\left(\Sigma^{-1} Q \Psi(\Lambda) Q^T\right) \quad \text{according to (A.10)}.
$$

Hence, $\Delta_{\theta,\Sigma}(\Psi)$ equals

$$
a_o^2 E_{\theta,\Sigma}\left[\text{tr}\left(\left\{\Psi(\Lambda)\right\}^2\right)\right] + 2a_o^2 E_{\theta,\Sigma}\left[\text{tr}\left(\Psi(\Lambda)\right)\right] - 2a_o E_{\theta,\Sigma}\left[\text{tr}\left(\Sigma^{-1} Q \Psi(\Lambda) Q^T\right)\right]. \tag{11}
$$

Then applying the Stein–Haff type identity in (A.4), with $\Phi = \Psi$, to the last expectation in (11) and expressing the various traces give rise to the unbiased estimator of the risk difference $\delta(\Psi)$ in (10), since $a_o = n \vee p$.

Finally, the improvement and dominance results are immediate. □

Remark 1. The loss function in (2) allows to get rid of the matrix Q in the expression of the risk of $\hat{\Sigma}_\Phi^{-1}$ through the fact that Q^- is the left inverse of Q and thanks to the Stein–Haff type identity in (A.4). Such a simplification does not occur with the usual quadratic loss $\mathrm{tr}[(\hat{\Sigma}^{-1} - \Sigma^{-1})^2]$ and with the data-based loss $\mathrm{tr}[(\hat{\Sigma}^{-1} - \Sigma^{-1})^2 S]$ used by Kubokawa and Srivastava [10].

2.2.1. A Haff type estimator. As an exemple, the estimator we consider in the following parallels the orthogonally invariant estimator used by Haff [8] in the case where S is invertible, and hence, is an extension involving Z. Let

$$\hat{\Sigma}^{-1}_{\Psi^{HF}} = a_o \left(S^+ + (Q^-)^T \Psi^{HF} Q^- \right) \quad \text{with} \quad \psi_i^{HF} = c \frac{\lambda_i^{-1}}{\mathrm{tr}(\Lambda^{-1})} \quad \text{and} \quad c > 0. \tag{12}$$

Using Theorem 1, we can see that $\hat{\Sigma}^{-1}_{\Psi^{HF}}$ improves over $\hat{\Sigma}^{-1}_o$ as soon as

$$0 < c \leq \frac{2(2k - n \wedge p - 1)}{n \vee p} \tag{13}$$

since $2k - n \wedge p - 1 \geq 0$. Indeed the unbiased estimator of the risk difference $\delta(\Psi)$ in (10) becomes

$$\begin{aligned} \delta(\Psi^{HF}) &= \sum_{i=1}^{k} \left\{ (n \vee p) \frac{\lambda_i^{-2}}{\mathrm{tr}^2\left(\Lambda^{-1}\right)} c^2 - 2(2k - n \wedge p - 1) \frac{\lambda_i^{-1}}{\mathrm{tr}(\Lambda^{-1})} c + 4 \left(\frac{\lambda_i^{-2}}{\mathrm{tr}^2\left(\Lambda^{-1}\right)} - \frac{\lambda_i^{-1}}{\mathrm{tr}\left(\Lambda^{-1}\right)} \right) c \right\} \\ &= (n \vee p) \frac{\mathrm{tr}(\Lambda^{-2})}{\mathrm{tr}^2\left(\Lambda^{-1}\right)} c^2 - 2(2k - n \wedge p - 1)c + 4 \left(\frac{\mathrm{tr}(\Lambda^{-2})}{\mathrm{tr}^2\left(\Lambda^{-1}\right)} - 1 \right) c. \end{aligned}$$

Using the fact that $\mathrm{tr}(\Lambda^{-2}) \leq \mathrm{tr}^2\left(\Lambda^{-1}\right)$, an upper bound for $\delta(\Psi^{HF})$ is given by

$$\delta(\Psi^{HF}) \leq (n \vee p) c^2 - 2(2k - n \wedge p - 1)c,$$

which is non positive since Condition (13) holds.

It is worth noting that, in the case where $m > p > n$, the improvement condition in (13) coincides with the one given in Kubokawa and Srivastava [10] for the orthogonally invariant Haff-type estimator. Actually, the roles of $(Q^-)^T$ and Λ^{-1} are respectively played by the eigenvectors matrix H and the eigenvalues L of S. This is due to the fact that the data-based loss (2) coincides with their loss function for the class of orthogonally invariant estimators (see Remark A.4).

2.2.2. A Stein type estimator. Let

$$\hat{\Sigma}^{-1}_{\Psi^{ST}} = a_o \left(S^+ + (Q^-)^T \Psi^{ST} Q^- \right) \quad \text{with} \quad \psi_i^{ST} = \frac{2i - n \wedge p - 1}{n \vee p} \tag{14}$$

for $i = 1, \ldots, k$. We will see that $\hat{\Sigma}^{-1}_{\Psi^{ST}}$ improves over $\hat{\Sigma}^{-1}_o$.

Applying Theorem 1, $\delta(\Psi)$ in (10) becomes

$$\delta(\Psi^{ST}) = \sum_{i=1}^{k} \left\{ (n \vee p)(\psi_i^{ST})^2 - 2(2k - n \wedge p - 1)\psi_i^{ST} + 4\lambda_i \frac{\partial \psi_i^{ST}}{\partial \lambda_i} + 4 \sum_{j>i} \frac{\lambda_i \psi_i^{ST} - \lambda_j \psi_j^{ST}}{\lambda_i - \lambda_j} \right\}.$$

Note that

$$\sum_{i=1}^{k} \sum_{j>i} \frac{\lambda_i \psi_i^{ST} - \lambda_j \psi_j^{ST}}{\lambda_i - \lambda_j} = \sum_{i=1}^{k} \sum_{j>i} \psi_i^{ST} + \sum_{i=1}^{k} \sum_{j>i} \lambda_i \frac{\psi_i^{ST} - \psi_j^{ST}}{\lambda_i - \lambda_j}.$$

Using the fact that, for any $j > i$, $\lambda_j < \lambda_i$, and $\psi_j^{ST} > \psi_i^{ST}$, we have

$$\sum_{i=1}^{k} \sum_{j>i} \lambda_i \frac{\psi_i^{ST} - \psi_j^{ST}}{\lambda_i - \lambda_j} \leq 0.$$

Then

$$\sum_{i=1}^{k} \sum_{j>i} \frac{\lambda_i \psi_i^{ST} - \lambda_j \psi_j^{ST}}{\lambda_i - \lambda_j} \leq \sum_{i=1}^{k} \sum_{j>i} \psi_i^{ST} = \sum_{i=1}^{k} (k - i) \psi_i^{ST}.$$

Therefore, according to (14), an upper bound of $\delta(\Psi^{ST})$ is given by

$$\delta(\Psi^{ST}) \le \sum_{i=1}^{k} \left\{ (n \vee p)(\psi_i^{ST})^2 - 2(2i - n \wedge p - 1)\psi_i^{ST} \right\} = -(n \vee p) \sum_{i=1}^{k} (\psi_i^{ST})^2,$$

which is non positive, and hence the result follows.

2.3. Truncated Estimators

In this subsection, we present new estimators which are truncated versions of the estimators considered in Subsection 2.2. This consists in replacing in (8) the function Ψ by $\Psi^{TR} = \text{diag}\left(\psi_i^{TR}\right)_{i=1,\dots,k}$ where, for any $i = 1, \dots, k$,

$$\psi_i^{TR} = \max\left\{ \psi_i, \frac{m + n \vee p}{a_o(1+\lambda_i)} - 1 \right\}.$$

Thus, the corresponding truncated estimators are of the form

$$(\hat{\Sigma}_{\Psi}^{-1})^{TR} = a_o \left(S^{+} + (Q^{-})^T \Psi^{TR} Q^{-} \right). \tag{15}$$

The following theorem yields a dominance result of the truncated estimators in (15) over the non truncated estimators in (8).

Theorem 2. *Let $\Psi(\Lambda) = \text{diag}\left(\psi_1(\Lambda), \dots, \psi_i(\Lambda), \dots, \psi_k(\Lambda)\right)$ be a $k \times k$ diagonal matrix function of Λ in (6) such that, for any $i = 1, \dots, k$, the function $\lambda_i \mapsto \psi_i\left(\lambda_1, \dots, \lambda_i, \dots, \lambda_k\right)$ is absolutely continuous and non negative.*

Any truncated estimator of the form $(\hat{\Sigma}_{\Psi}^{-1})^{TR}$ in (15) dominates the non-truncated estimator $\hat{\Sigma}_{\Psi}^{-1}$ in (8), provided that $Pr(\Psi^{TR} \neq \Psi) > 0$.

Proof. According to (11), the risk difference between $\hat{\Sigma}_{\Psi^{TR}}^{-1}$ in (15) and $\hat{\Sigma}_{\Psi}^{-1}$ in (8) is expressed as

$$\begin{aligned} \Delta(\Psi^{TR}) = {} & a_o^2 E_{\theta,\Sigma}\left[\text{tr}((\Psi^{TR})^2 - \Psi^2)\right] + 2a_o^2 E_{\theta,\Sigma}\left[\text{tr}(\Psi^{TR} - \Psi)\right] \\ & - 2a_o E_{\theta,\Sigma}\left[\text{tr}(\Sigma^{-1} Q (\Psi^{TR} - \Psi) Q^T)\right], \end{aligned}$$

where the diagonal elements of $\Psi^{TR} - \Psi$ are non-negative. Applying Lemma A.2 with $\Phi = \Psi^{TR} - \Psi$, an upper bound for the risk difference is given by

$$\begin{aligned} \Delta(\Psi^{TR}) \le {} & 2a_o^2 E_{\theta,\Sigma}\left[\text{tr}(\Psi^{TR} - \Psi)\right] + a_o^2 E_{\theta,\Sigma}\left[\text{tr}((\Psi^{TR})^2 - \Psi^2)\right] \\ & - 2a_o (m + n \vee p) E_{\theta,\Sigma}\left[\text{tr}(\left\{\Psi^{TR} - \Psi\right\}(I_k + \Lambda)^{-1})\right] \\ = {} & a_o^2 E_{\theta,\Sigma}\left[\sum_{i=1}^{k} \left\{ (\psi_i^{TR} - \psi_i)\left(\psi_i^{TR} + \psi_i - 2\left(\frac{m + n \vee p}{a_o(1+\lambda_i)} - 1 \right) \right) \right\} \right]. \end{aligned}$$

Let I the sub-set of $\{1, \dots, k\}$ such that $\psi_i^{TR} \neq \psi_i$, and hence, such that

$$\psi_i^{TR} = \frac{m + n \vee p}{a_o(1+\lambda_i)} - 1.$$

By assumption, $I \neq \emptyset$ with positive probability. Therefore

$$\Delta(\Psi^{TR}) \le a_o^2 E_{\theta,\Sigma}\left[\sum_{i \in I} -\left(\psi_i^{TR} - \psi_i\right)^2 \right] < 0,$$

which guarantees domination of the truncated estimators in (15) over the non-truncated estimators in (8). □

Remark 2. Continuing Remark 1, the choice of the loss function (2) works in favor of our group of truncated estimators in (15) in the sense that, thanks to the presence of the statistics S, it allows to highlight that these estimators clearly improve over the usual estimators aS^{+}.

Theorem 2 applies to the examples given in Subsection 2.2. Thus, the Haff type estimator in (12) is dominated by the truncated Haff type estimator

$$(\hat{\Sigma}^{-1}_{\Psi^{HF}})^{TR} = a_o \left(S^+ + (Q^-)^T \left(\Psi^{HF}\right)^{TR} Q^- \right) \tag{16}$$

where

$$\left(\Psi^{HF}\right)^{TR} = \operatorname{diag}\left(\max\left\{ c\frac{\lambda_i^{-1}}{\operatorname{tr}(\Lambda^{-1})}, \frac{m + n \vee p}{a_o(1+\lambda_i)} - 1 \right\} \right)_{i=1,\dots,k}.$$

As for the Stein-type estimator in (14), it is dominated by the truncated Stein-type estimator

$$(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{TR} = a_o \left(S^+ + (Q^-)^T \left(\Psi^{ST}\right)^{TR} Q^- \right) \tag{17}$$

where

$$\left(\Psi^{ST}\right)^{TR} = \operatorname{diag}\left(\max\left\{ \frac{2i - n \wedge p - 1}{n \vee p}, \frac{m + n \vee p}{a_o(1+\lambda_i)} - 1 \right\} \right)_{i=1,\dots,k}.$$

3. NUMERICAL STUDY

In this section, although we formally proved in Subsection 2.3 that the non-truncated estimators are dominated by the truncated estimators, we illustrate numerically this performance. While we do not have a theoretical result, it is also interesting to compare the truncated estimators with the orthogonally invariant estimators.

To this end, we consider the following structures of the scatter matrix Σ: the identity matrix $\Sigma_1 = I_p$ and an autoregressive structure Σ_2 with coefficient 0.9 (i.e., a $p \times p$ matrix where the (i, j)th element is $0.9^{|i-j|}$). As for the mean matrix θ, we deal with: a random matrix θ_1 where the (i, j)th element is generated from a uniform distribution over $[0, 1)$ and the null matrix $\theta_2 = 0_{m\times p}$. To assess how an alternative estimator $\hat{\Sigma}^{-1}$ improves over $\hat{\Sigma}_o^{-1}$, we compute the percentage reduction in average loss (PRIAL), defined as

$$\text{PRIAL}(\hat{\Sigma}^{-1}) = \frac{\text{average loss of } \hat{\Sigma}_o^{-1} - \text{average loss of } \hat{\Sigma}^{-1}}{\text{average loss of } \hat{\Sigma}_o^{-1}},$$

based on independent Monte-Carlo replications.

3.1. Comparison of Truncated and Non-Truncated Estimators

We compare here the performance of the following estimators:

- $\hat{\Sigma}^{-1}_{\Psi_{HF}}$ given in (12) with $c = 2(2k - n \wedge p - 1)/(n \vee p)$;
- $(\hat{\Sigma}^{-1}_{\Psi_{HF}})^{TR}$ given in (16);
- $\hat{\Sigma}^{-1}_{\Psi_{ST}}$ given in (14);
- $(\hat{\Sigma}^{-1}_{\Psi_{ST}})^{TR}$ given in (17).

Note that the value of c we consider is the upper bound in (13) which turns out to provide higher Prial's in the range of the values of c allowed by (13). This choice is all the more legitimate as we have observed, in other simulations we have not reproduced here, that there exists a range of values of c, larger than the one that our theory provides, for which $\hat{\Sigma}^{-1}_{\Psi_{HF}}$ improves over $\hat{\Sigma}_o^{-1}$.

Table 1 shows the results based on 10 000 independent Monte-Carlo replications for all possible pairs of Σ and θ structures. The values of m, n, and p are chosen from all possible combinaisons of 15, 30, and 45.

Table 1. Prial's (%) of the proposed estimators based on 10 000 independent Monte–Carlo replications for any ordering of m, n, and p

Σ	θ	m	n	p	$\hat{\Sigma}^{-1}_{\Psi^{HF}}$	$\hat{\Sigma}^{-1}_{\Psi^{ST}}$	$(\hat{\Sigma}^{-1}_{\Psi^{HF}})^{TR}$	$(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{TR}$
Σ_1	θ_1	30	15	45	6.921	27.603	17.134	31.538
		15	30	45	−0.008	14.47	0.355	15.265
		15	45	30	−0.008	14.408	0.348	15.194
		45	15	30	10.669	41.99	29.387	49.847
		45	30	15	10.659	41.487	29.172	49.204
		30	45	15	6.928	27.539	17.142	31.463
	θ_2	30	15	45	6.938	29.557	17.747	33.839
		15	30	45	−0.007	14.835	0.363	15.682
		15	45	30	−0.008	14.791	0.374	15.647
		45	15	30	10.739	44.31	30.513	52.758
		45	30	15	10.739	44.373	30.462	52.78
		30	45	15	6.92	29.564	17.65	33.793
Σ_2	θ_1	30	15	45	6.802	24.836	13.381	26.672
		15	30	45	−0.008	14.269	0.321	14.928
		15	45	30	−0.008	14.436	0.355	15.229
		45	15	30	10.554	39.147	24.985	44.361
		45	30	15	10.706	41.212	29.181	48.953
		30	45	15	6.902	27.086	16.992	30.93
	θ_2	30	15	45	6.86	25.971	14.081	28.091
		15	30	45	−0.008	14.665	0.345	15.416
		15	45	30	−0.008	14.796	0.365	15.642
		45	15	30	10.654	40.822	26.553	46.868
		45	30	15	10.703	44.469	30.268	52.723
		30	45	15	6.973	29.591	17.881	33.941

The results reveal that, on the whole, the behavior of the estimators does not change significantly with respect to the structures of Σ and θ and, for any pair, the best performances are obtained in the case where $m > n \wedge p$. Note that the truncated estimators dominate the non-truncated estimators, the truncated Stein type estimator $(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{TR}$ $(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{TR}$ outperforming the others with Prial's that can reach 52%. Note also that, in the case where $2m - n \wedge p - 1 < 0$ so that condition (13) cannot be satisfied, the Haff type estimator $\hat{\Sigma}^{-1}_{\Psi^{HF}}$ fails to dominate the usual estimator $\hat{\Sigma}^{-1}_{o}$. For example, when $m = 15$, $n = 30$, and $p = 45$, the Prial's are negative.

3.2. Comparison of Truncated and Orthogonally Invariant Estimators

We compare numerically the performance of the truncated estimators $(\hat{\Sigma}^{-1}_{\Psi^{HF}})^{TR}$ and $(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{TR}$ with their respectively orthogonally invariant versions

$$(\hat{\Sigma}^{-1}_{\Psi^{HF}})^{OI} = a_o\left(S^+ + H\Psi^{HF}H^T\right) \quad \text{with} \quad \psi_i^{HF} = c\frac{l_i}{\text{tr}(L)} \quad \text{for} \quad c = \frac{2(n \wedge p - 1)}{n \vee p}$$

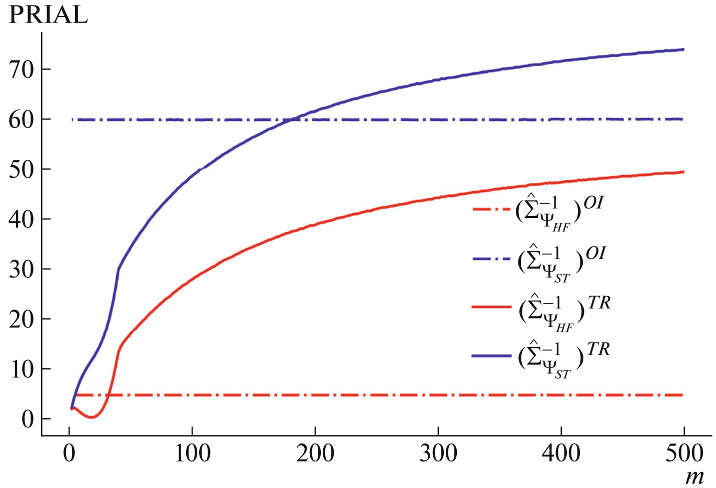

Fig. 1. Effect of m on the truncated and the orthogonally invariant estimators for $(n, p) = (40, 80)$. The structures of Σ and θ are respectively Σ_1 and θ_1.

and

$$(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{OI} = a_o \left(S^+ + H\Psi^{ST}H^T\right) \quad \text{with} \quad \psi_i^{ST} = \frac{2i - n \wedge p - 1}{n \vee p},$$

where $i = 1, \ldots, n \wedge p$ and where H and L are defined at the beginning of Subsection 2.1.

Note that analytical dominance results of such estimators over $\hat{\Sigma}_o^{-1}$ have been given by many authors (see Remark A.4). In this study, we compare their numerical performance with the truncated estimators (16) and (17).

Figure 1 shows the Prial's of the estimators $(\hat{\Sigma}^{-1}_{\Psi^{HF}})^{TR}$ and $(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{TR}$ and their respective corresponding orthogonally invariant estimators $(\hat{\Sigma}^{-1}_{\Psi^{HF}})^{OI}$ and $(\hat{\Sigma}^{-1}_{\Psi^{ST}})^{OI}$ based on 5000 independent Monte-Carlo replications. We consider the structures Σ_1 and θ_1 for Σ and θ where the couple (n, p) is fixed to $(40, 80)$ and $m = 1, \ldots, 500$. Thus we encounter tricky cases since the sample covariance matrix S is non-invertible ($n < p$).

The important point to note here is that the risks of a truncated estimator (continuous lines) and of its corresponding orthogonally invariant (discountinuous lines) intersect, these last estimators presenting an almost constant risk. Also, the performance of both truncated estimators increases as m increases with trend changing around $m = n = 40$, the Stein's truncated estimator outperforming the others from $m = 180$.

4. CONCLUSIONS AND PERSPECTIVES

In this paper, we proposed an improved class of a non-truncated estimators of the scatter matrix Σ^{-1} depending not only on the sample covariance matrix S, such as orthogonally invariant estimators, but also on the information contained in the sample mean Z. This allowed us to construct truncated estimators that outperform these non-truncated estimators. Numerical results show that the risks of the truncated estimators and their corresponding orthogonally invariant estimators intersect and that the truncated estimators outperform the orthogonally invariant estimators when $m \to \infty$. This result needs to be proved analytically. It would be of interest to extend the results of this paper to the case of normal mixture distributions with respect to Σ and, more generally, to the case of elliptically symmetric distributions.

APPENDIX A

We provide here materials for the proofs of Theorems 1 and 2 where we place ourselves in the context of Lemma A.3 at the end of this appendix.

A.1. A Stein–Haff Identity

A key tool for the proof of these theorems is the Stein–Haff identity given in the following lemma.

Lemma A.1 (Equation (4.4) of Tsukuma and Kubokawa [17]). *Let Z and U two random matrices with respective dimension $m \times p$ and an $n \times p$ such that $(Z^T, U^T)^T$ has joint density*

$$(2\pi)^{-(m+n)p/2}|\Sigma|^{-(m+n)/2}\exp[-\mathrm{tr}\{(z-\theta)\Sigma^{-1}(z-\theta)^T + u\Sigma^{-1}u^T\}/2]. \tag{A.1}$$

Let Q and Λ as in (A.7) *and let also $\Phi(\Lambda) = \mathrm{diag}\,(\phi_1(\Lambda), \ldots, \phi_i(\Lambda), \ldots, \phi_k(\Lambda))$ be a $k \times k$ diagonal matrix such that, for any $i = 1, \ldots, k$, the function $\lambda_i \mapsto \phi_i(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_k)$ is absolutely continuous.*

Assuming that $E_{\theta,\Sigma}\big[|tr(\Sigma^{-1}Q\Phi Q^\top)|\big] < \infty$, where $E_{\theta,\Sigma}$ denotes the expectation with respect to (A.1), *we have*

$$E_{\theta,\Sigma}\left[\mathrm{tr}(\Sigma^{-1}Q\Phi Q^\top)\right] = E_{\theta,\Sigma}\left[\sum_{i=1}^{k}\left\{\alpha_i\phi_i - 2\lambda_i\frac{\partial\phi}{\partial\lambda_i} - 2\sum_{j>i}^{k}\frac{\phi_i-\phi_j}{\lambda_i-\lambda_j}\lambda_j\right\}\right], \tag{A.2}$$

where

$$\forall i = 1, \ldots, k \quad \alpha_i = |n-p| + 2i - 1.$$

Remark A.3. Throughout this paper, we use the modified version of identity (A.2) given in (A.4) below. This identity is established as follows. As for $j > i = 1, \ldots, k$ we have

$$\frac{\phi_i-\phi_j}{\lambda_i-\lambda_j}\lambda_j = \frac{\phi_i\lambda_i-\phi_j\lambda_j}{\lambda_i-\lambda_j} - \phi_i$$

and

$$\sum_{j>i}^{k}\frac{\phi_i-\phi_j}{\lambda_i-\lambda_j}\lambda_j = \sum_{j>i}^{k}\frac{\phi_i\lambda_i-\phi_j\lambda_j}{\lambda_i-\lambda_j} - (k-i)\phi_i \tag{A.3}$$

since

$$\sum_{j>i}^{k}\phi_i = (k-i)\phi_i.$$

Then substituting (A.3) for (A.2) gives

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}Q\Phi Q^T\right)\right]$$

$$= E_{\theta,\Sigma}\left[\sum_{i=1}^{k}\left\{(n\vee p - n\wedge p + 2k - 1)\phi_i - 2\lambda_i\frac{\partial\phi_i}{\partial\lambda_i} - 2\sum_{j>i}\frac{\lambda_i\phi_i-\lambda_j\phi_j}{\lambda_i-\lambda_j}\right\}\right] \tag{A.4}$$

since

$$\alpha_i + 2(k-i) = n\vee p - n\wedge p + 2k - 1.$$

A.2. The Case where $k = n \wedge p$

Note that, as mentioned at the beginning of Subsection 2.2, when $k = n \wedge p$, the estimators in (8) can be rewritten as $\hat{\Sigma}_\Phi^{-1} = (Q^-)^T\Phi Q^-)$ where $\Phi = a_o\,(I_k + \Psi)$. Although not used in the rest of the article, for completeness, we give the risk of such estimators in the following proposition.

Proposition A.1. *Let Q and Λ as in* (A.7) *below and let $\Phi = \mathrm{diag}\big(\phi_1(\Lambda), \ldots, \phi_i(\Lambda), \ldots, \phi_k(\Lambda)\big)$ be a $k \times k$ diagonal matrix such that, for any $i = 1, \ldots, k$, the function $\lambda_i \mapsto \phi_i(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_k)$ is absolutely continuous and non-negative.*

The risk function in (3) *of the estimator* $\hat{\Sigma}_\Phi^{-1} = (Q^-)^T \Phi Q^-$ *is given by*

$$R\Big(\Sigma^{-1}, \hat{\Sigma}_\Phi^{-1}\Big) = E_{\theta,\Sigma}\left[\sum_{i=1}^{k}\left\{\phi_i^2 - 2(n \vee p - n \wedge p + 2k - 1)\phi_i + 4\lambda_i \frac{\partial \phi_i}{\partial \lambda_i} + 4\sum_{j>i}\frac{\lambda_i\phi_i - \lambda_j\phi_j}{\lambda_i - \lambda_j}\right\}\right]$$
$$+ E_{\theta,\Sigma}\Big[\mathrm{tr}\Big(\Sigma^{-1}S\Big)^2\Big].$$

Remark A.4. When $m > p > n$, for the class of orthogonally invariant estimators, the risk expression in Proposition A.1 parallels the one associated to the data-based loss $\mathrm{tr}[(\hat{\Sigma}^{-1} - \Sigma^{-1})^2 S^2]$ provided by Kubokawa and Srivastava [10] in their Proposition 2.12 where the role of the λ_i's is played by l_i^{-1}'s. They showed that these estimators, that depend only on S, improve over the optimal estimator in (4).

Proof. According to the loss function (2), the risk function of any estimators of the form $\hat{\Sigma}_\Phi^{-1} = (Q^-)^T \Phi Q^-$ is expressed as

$$R\left(\Sigma^{-1}, \hat{\Sigma}_\Phi^{-1}\right) = E_{\theta,\Sigma}\left[\mathrm{tr}\left(\hat{\Sigma}_\Phi^{-1}S\right)^2\right] - 2E_{\theta,\Sigma}\left[\Sigma^{-1}S\hat{\Sigma}_\Phi^{-1}S\right] + E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}S\right)^2\right]$$
$$= E_{\theta,\Sigma}\left[\mathrm{tr}\left((Q^-)^T\Phi Q^- S\right)^2\right] + E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}S\right)^2\right] - 2E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}S(Q^-)^T\Phi Q^- S\right)\right].$$

Using the fact that $S(Q^-)^T = Q$ and that $Q^-Q = I_k$, we have

$$R\left(\Sigma^{-1}, \hat{\Sigma}_\Phi^{-1}\right) = E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Phi\right)^2\right] - 2E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}Q\Phi Q^T\right)\right] + E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}S\right)^2\right].$$

Applying the Stein–Haff type identity in (A.4) to the second term in the right-hand side of the last equality gives the desired result. □

Remark A.5. The loss function in (2) allows to get rid of the matrix Q in the expression of the risk of $\hat{\Sigma}_\Phi^{-1}$ through the fact that Q^- is the left inverse of Q and thanks to the Stein–Haff type identity in Proposition A.4. Such a simplification does not occur with the usual quadratic loss $\mathrm{tr}[(\hat{\Sigma}^{-1} - \Sigma^{-1})^2]$ and with the data-based losses $\mathrm{tr}[(\hat{\Sigma}^{-1} - \Sigma^{-1})^2 S^r]$ used by Kubokawa and Srivastava [10] for $r = 1, 2$. The choice of our loss function works in favor of our group of truncated estimators in (15) in the sense that, thanks to the presence of the statistics S, it allows to highlight that these estimators clearly improve over the usual estimators aS.

A.3. Determination of the Optimal Constant "a_o"

Here, we prove the statement in (4). The risk of aS^+ equals

$$R\left(\Sigma^{-1}, aS^+\right) = E_{\theta,\Sigma}\left[\mathrm{tr}\left[\left(\hat{\Sigma}_o^{-1} - \Sigma^{-1}\right)S\right]^2\right]$$
$$= a^2 E_{\theta,\Sigma}\left[\mathrm{tr}\big(S^+SS^+S\big)\right] - 2aE_{\theta,\Sigma}\left[\mathrm{tr}\big(\Sigma^{-1}SS^+S\big)\right] + E_{\theta,\Sigma}\left[\mathrm{tr}\Big(\Sigma^{-1}S\Big)^2\right]$$
$$= a^2(n \wedge p) - 2aE_{\theta,\Sigma}\left[\mathrm{tr}\big(\Sigma^{-1}S\big)\right] + E_{\theta,\Sigma}\left[\mathrm{tr}\Big(\Sigma^{-1}S\Big)^2\right],$$

where we used the fact that $S^+SS^+S = S^+S$, $\mathrm{tr}\big(S^+S\big) = n \wedge p$ and $SS^+S = S$. Clearly, this polynomial in a is minimized for

$$a = a_o = \frac{E_{\theta,\Sigma}\left[\mathrm{tr}\big(\Sigma^{-1}S\big)\right]}{n \wedge p}. \tag{A.5}$$

Note that this expression does not depend on θ or Σ. Indeed, according to Lemma 2.1 of Haddouche et al. [7], if $G(z,s)$ be a $p \times p$ matrix function such that, for any fixed $z \in \mathbb{R}^{m\times p}$, $G(z,s)$ is weakly differentiable with respect to $s \in \mathbb{R}^{p\times p}$ and such that $E_{\theta,\Sigma}\big[|\mathrm{tr}(\Sigma^{-1}G(Z,S))|\big] < \infty$, we have

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^+G(Z,S)\right)\right]$$

$$= E_{\theta,\Sigma}\Big[\operatorname{tr}\Big(2SS^{+}\mathcal{D}_s\{SS^{+}G(Z,S)\}^T + (n - n\wedge p - 1)S^{+}G(Z,S)\Big)\Big], \tag{A.6}$$

where $\mathcal{D}_s\{\cdot\}$ is the Haff operator whose generic element is $\frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial S_{ij}}$, with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Using (A.6) with $G(Z,S) = S$ it follows that

$$E_{\theta,\Sigma}\left[\operatorname{tr}(\Sigma^{-1}S)\right] = E_{\theta,\Sigma}\left[\operatorname{tr}(\Sigma^{-1}SS^{+}S)\right]$$
$$= E_{\theta,\Sigma}\left[\operatorname{tr}\left(2SS^{+}\mathcal{D}_s\{S\} + (n - n\wedge p - 1)S^{+}S\right)\right].$$

Now, according to Lemma A.6 of Haddouche et al. [7], we have

$$2SS^{+}\mathcal{D}_s\{S\} = (p+1)SS^{+},$$

so that, as $\operatorname{tr}(SS^{+}) = n \wedge p$, it follows that

$$E_{\theta,\Sigma}\left[\operatorname{tr}(\Sigma^{-1}S)\right] = (p+1)(n\wedge p) + (n - n\wedge p - 1)(n\wedge p) = (n\vee p)(n\wedge p).$$

Hence, according to (A.5), $a_0 = n \vee p$.

Remark A.6. The optimal constant $a_o = n \vee p$ with respect to our data-based loss and the one in Kubokawa and Srivastava [10] are identical when $p > n$. This is not surprising since theses data-based losses coincide for the class of orthogonally invariant estimators as $\hat{\Sigma}_a^{-1}$ as mentioned in Remark 2.

A.4. A Fundamental Inequality

The following lemma is a key tool for the proof of Theorem 2.

Lemma A.2 (Theorem 3.1 of Tsukuma and Kubokawa [17]). *Let Q and Λ as in* (A.7) *below. Let also $\Phi = \operatorname{diag}(\phi_1(\Lambda), \ldots, \phi_i(\Lambda), \ldots, \phi_k(\Lambda))$ be a $k\times k$ diagonal matrix such that, for any $i = 1, \ldots, k$, the function $\lambda_i \mapsto \phi_i(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_k)$ is absolutely continuous and non negative. Then*

$$E_{\theta,\Sigma}\left[\operatorname{tr}(\Sigma^{-1}Q\Phi Q^T)\right] \geq (m + n\vee p)E_{\theta,\Sigma}\left[\operatorname{tr}(\Phi(I_k+\Lambda)^{-1})\right].$$

A.5. A Simultaneous Diagonalization Result

Here, we give a simultaneous diagonalization lemma of two matrices in the context of (5) and (7).

Lemma A.3 (Simultaneous diagonalization). *For Z and U two matrices with dimension $m\times p$ and $n\times p$, respectively, let $S = U^TU$ with* $\operatorname{rank}(S) = n\wedge p$ *and $W = Z^TZ$ with* $\operatorname{rank}(W) = m\wedge p$. *Let $S = HLH^T$ the eigenvalue decomposition of S where $L = \operatorname{diag}(l_1 > \cdots > l_i > \cdots > l_{n\wedge p} > 0)$ is the $(n\wedge p)\times(n\wedge p)$ diagonal matrix of the positive eigenvalues of S and H the $p\times(n\wedge p)$ semi-orthogonal matrix ($H^TH = I_{n\wedge p}$) of the corresponding eigenvectors. Let also $k = n\wedge m\wedge p$.*

Then there exists a $k\times p$ matrix Q^- such that the following simultaneous diagonalization of W and S holds

$$Q^-W(Q^-)^T = \Lambda \quad \text{and} \quad Q^-S(Q^-)^T = I_k, \tag{A.7}$$

where the $k\times k$ diagonal matrix Λ intervenes in the thin singular value decomposition of $ZHL^{-1/2}$, that is,

$$ZHL^{-1/2} = R\Lambda^{1/2}O^T \tag{A.8}$$

with

$$\Lambda = \operatorname{diag}(\lambda_1 > \cdots > \lambda_i > \cdots > \lambda_k > 0),$$

O an $(n\wedge p)\times k$ semi-orthogonal matrix ($O^TO = I_k$) and R an $m\times k$ semi-orthogonal matrix ($R^TR = I_k$). More precisely, we have $Q^- = O^TL^{-1/2}H^T$.

Besides, Q^- is the reflexive generalized inverse[1] *of $Q = HL^{1/2}O$, which are respectively $k \times p$ and $p \times k$ matrices. Also, Q^- is a left inverse of Q, that is,*

$$Q^- Q = I_k. \tag{A.9}$$

In addition, we have

$$S(Q^-)^T = Q \quad \textit{and} \quad Q^- S = Q^T, \tag{A.10}$$

$$SS^+ Q = Q \tag{A.11}$$

and

$$Q^T S^+ Q = I_k, \tag{A.12}$$

where $S^+ = HL^{-1}H^T$ is the Moore–Penrose inverse of S.

Proof. Through the expressions of Q^- and W we have

$$\begin{aligned} Q^- W (Q^-)^T &= O^T L^{-1/2} H^T Z^T Z H L^{-1/2} O \\ &= O^T O \Lambda^{1/2} R^T R \Lambda^{1/2} O^T O \quad \text{according to (A.8)} \\ &= \Lambda \quad \text{thanks to the semi-orthogonality of } R \text{ and } O. \end{aligned}$$

This is the first diagonalization in (A.7). As for the second diagonalization in (A.7), we can write, according to the expression of Q^- and to the eigenvalue decomposition of S,

$$Q^- S (Q^-)^T = O^T L^{-1/2} H^T H L H^T H L^{-1/2} O = I_k,$$

thanks to the semi-orthogonality of H and O.

Now, with the same semi-orthogonality arguments, (A.9) follows immediately from

$$Q^- Q = O^T L^{-1/2} H^T H L^{1/2} O = I_k$$

and (A.10) from

$$S(Q^-)^T = HLH^T H L^{-1/2} O = HL^{1/2} O = Q$$

and

$$Q^- S = O^T L^{-1/2} H^T H L H^T = O^T L^{1/2} H^T = Q^T.$$

Finally, Equalities (A.11) and (A.12) are obtained as follows

$$SS^+ Q = HLH^T H L^{-1} H^T H L^{1/2} O = HL^{1/2} O = Q$$

and

$$Q^T S^+ Q = O^T L^{1/2} H^T H L^{-1} H L^{1/2} O = I_k.$$

□

ACKNOWLEDGMENTS

We thank Pr. Stéphane Canu (INSA de Rouen Normandie) for helpful discussions about the proof of Lemma A.3. We are grateful to two anonymous referees for their careful reading which allowed us to write an improved version of this paper.

FUNDING

This work was supported by ongoing institutional funding. No additional grants to carry out or direct this particular research were obtained.

[1] This means that Q^- is a generalized inverse of Q, that is, $QQ^-Q = Q$ and that Q is a generalized inverse of Q^-, that is, $Q^-QQ^- = Q^-$. This can be checked thanks to the semi-orthogonality on O and H: $QQ^-Q = HL^{1/2}O\, O^T L^{-1/2} H^T HL^{1/2}O = Q$ and $Q^-QQ^- = O^T L^{-1/2} H^T\, HL^{1/2}O\, O^T L^{-1/2} H^T = Q^-$.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. D. Boukehil, D. Fourdrinier, F. Mezoued, and W. E. Strawderman, “Estimation of the inverse scatter matrix for a scale mixture of Wishart matrices under Efron-Morris type losses,” J. Statist. Plann. Inference **215**, 368–387 (2021).
2. S. Canu and D. Fourdrinier, “Unbiased risk estimates for matrix estimation in the elliptical case,” J. Multivar. Anal. **158**, 60–72 (2017).
3. D. Fourdrinier and W. E. Strawderman, “Robust minimax Stein estimation under invariant data-based loss for spherically and elliptically symmetric distributions,” Metrika **78** (4), 461–484 (2015).
4. D. Fourdrinier, F. Mezoued, and M. T. Wells, “Estimation of the inverse scatter matrix of an elliptically symmetric distribution,” J. Multivar. Anal. **143**, 32–55 (2016).
5. D. Fourdrinier, A. M. Haddouche, and F. Mezoued, “Covariance matrix estimation under data-based loss,” Statistics and Probability Letters **177**, 109160 (2021).
6. G. Golub and C. van Loan, *Matrix Computations* (JHU Press, 3rd ed., 1996).
7. A. M. Haddouche, D. Fourdrinier, and F. Mezoued, “Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions,” J. Multivar. Anal. **181**, 104680 (2021).
8. L. R. Haff, “Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity,” Ann. Statist. **7** (6), 1264–1276, 11 (1979).
9. T. Kubokawa and M. Srivastava, “Estimating the covariance matrix: A new approach,” J. Multivar. Anal. **86** (1), 28–47 (2003).
10. T. Kubokawa and M. Srivastava, “Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data,” J. Multivar. Anal. **99** (9), 1906–1928 (2008).
11. T. Kubokawa and M. Tsai, “Estimation of covariance matrices in fixed and mixed effects linear models,” J. Multivar. Anal. **97** (10), 2242–2261 (2006).
12. B. K. Sinha and M. Ghosh, “Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss,” Statistics and Decisions **5** (3–4), 201–228 (1987).
    https://doi.org/10.1524/strm.1987.5.34.201
13. M. S. Srivastava, “Singular Wishart and multivariate Beta distributions,” Ann. Statis. **31** (5), 1537–1560 (2003).
14. C. Stein, “Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean,” Ann. Inst. Stat. Math. **16** (1), 155–160 (1964).
15. A. Takemura, “An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population,” Tsukuba J. Math. **8**, 367–376 (1984).
16. H. Tsukuma and T. Kubokawa, “A unified approach to estimating a normal mean matrix in high and low dimensions,” J. Multivar. Anal. **139**, 312–328 (2015).
17. H. Tsukuma and T. Kubokawa, “Unified improvements in estimation of a normal covariance matrix in high and low dimensions,” J. Multivar. Anal. **143**, 233–248 (2016).
18. H. Tsukuma and T. Kubokawa, “Multivariate linear model and group invariance,” in: *Shrinkage Estimation for Mean and Covariance Matrices* (Springer, 2020), pp. 27–33.