INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus
Mathématique

Anis M. Haddouche and Wei Lu

A unified approach for covariance matrix estimation under Stein loss

Volume 360 (2022), p. 1093-1098

https://doi.org/10.5802/crmath.356

Les Comptes Rendus. Mathématique *sont membres du*
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 1778-3569

Comptes Rendus
Mathématique
2022, Vol. 360, p. 1093-1098
https://doi.org/10.5802/crmath.356

Statistics / *Statistiques*

A unified approach for covariance matrix estimation under Stein loss

Anis M. Haddouche *, [a] and Wei Lu [a]

[a] INSA Rouen, Normandie Univ, UNIROUEN, UNIHAVRE, LITIS and LMI, avenue de l'Université, BP 8, Saint-Étienne-du-Rouvray, 76801, France

E-mails: mohamed.haddouche@insa-rouen.fr (A. M. Haddouche), wei.lu@insa-rouen.fr (W. Lu)

Abstract. In this paper, we address the problem of estimating a covariance matrix of a multivariate Gaussian distribution, from a decision theoretic point of view, relative to a Stein type loss function. We investigate the case where the covariance matrix is invertible and the case when it is non–invertible in a unified approach.

Résumé. Dans cet article, nous abordons le problème de l'estimation d'une matrice de covariance d'une distribution gaussienne multivariée, du point de vue de la théorie de la décision, par rapport à une fonction de coût de type Stein. Nous étudions dans une approche unifiée le cas où la matrice de covariance est inversible et le cas où elle n'est pas inversible.

Mathematical subject classification (2010). 62H12, 62F10, 62C99.

Manuscript received 13 July 2021, revised 4 January 2022, accepted 13 March 2022.

1. Introduction

Let $\boldsymbol{X}$ be an observed $p \times n$ matrix of the form

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{Z}, \tag{1}$$

where $\boldsymbol{B}$ is a $p \times r$ matrix of unknown parameters, with $r \leq p$, and $\boldsymbol{Z}$ is a $r \times n$ random matrix. Assume that r is known and that the columns of $\boldsymbol{Z}$ are identically and independently distributed as the r-dimensional multivariate normal distribution $\mathcal{N}_r(0_r, \boldsymbol{I}_r)$. Then the columns of $\boldsymbol{X}$ are identically and independently distributed from the p-dimensional multivariate normal $\mathcal{N}_p(0_p, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^T$ is the unknown $p \times p$ covariance matrix with

$$\operatorname{rank}(\boldsymbol{\Sigma}) = r \leq p\,.$$

It follows that the $p \times p$ sample covariance matrix $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^T$ has a singular Wishart distribution (see Srivastava [9]) such that

$$\operatorname{rank}(\boldsymbol{S}) = \min(n, r) = q \leq p\,,$$

* Corresponding author.

with probability one. Note that, as the matrices $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ are non–invertible, we denote in the following by $\boldsymbol{S}^+$ and $\boldsymbol{\Sigma}^+$ the Moore-Penrose inverses of $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ respectively.

We consider the problem of estimating the covariance matrix $\boldsymbol{\Sigma}$ under the Stein type loss function

$$L\left(\widehat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}}\right) - \ln\left|\Lambda\left(\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}}\right)\right| - q, \tag{2}$$

where $\widehat{\boldsymbol{\Sigma}}$ estimates $\boldsymbol{\Sigma}$ and $\Lambda(\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}})$ is the diagonal matrix of the q positives eigenvalues of $\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}}$. Note that this loss function is based on the Kullback–Leibler divergence between two multivariate normal distributions with the same means and with covariances $\widehat{\Sigma}$ and Σ. It is an extension of the usual Stein loss function (see Stein [10]). Thus, the corresponding risk function is given by

$$R\left(\widehat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}\right) = E\left[L\left(\widehat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}\right)\right],$$

where $E(\cdot)$ denotes the expectation with respect to the model (1). Note that the loss function (2) is an adaptation of the original Stein loss function (see Stein [11]) to the context of the model (1) (see Tsukuma [13] for more details).

As related by Ledoit and Wolf [8], the difficulty of covariance estimation is commonly characterized by the ratio p/n. In fact, the usual estimators of the form

$$\widehat{\boldsymbol{\Sigma}}_a = a\boldsymbol{S} \quad \text{with} \quad a > 0, \tag{3}$$

perform poorly when $n, p \to \infty$ with $p/n \to c > 0$. Indeed, the larger (smaller) eigenvalues of $\boldsymbol{\Sigma}$ are overestimated (underestimated) by those estimators. Hence, in this asymptotic context, alternative estimators dominating (with lower risk functions) these usual estimators are needed. A possible approach is to regularize the eigenvalues of $\widehat{\boldsymbol{\Sigma}}_a$, which gives rise to the class of orthogonally invariant estimators (see Takemura [12]) as those in (5) below.

Note that, as we consider the model (1), we deal in a unified approach, with the following cases.

(i) $n < r = p$: $\boldsymbol{\Sigma}$ is invertible of rank p and $\boldsymbol{S}$ is non–invertible of rank n;
(ii) $r = p \le n$: $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are invertible;
(iii) $r < p \le n$: $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are non–invertible of rank r;
(iv) $r \le n < p$: $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are non–invertible of rank r;
(v) $n < r < p$: $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are non–invertible of ranks r and n respectively.

The class of orthogonally invariant estimators was considered by various authors. For instance, see Stein [11], Dey and Srivastava [2] and Haff [5] for the case (i), Konno [6] and Haddouche et al. [4] for the cases (i) and (ii). See also Chételat and Wells [1] for the cases (iii) and (iv). Recently, Tsukuma and Kubokawa [13] extended the Stein [11] estimator to the five possible cases, cited above, in a unified approach. Similarly, we extend here the class of Haff [5] estimators to the context of the model (1).

The rest of this paper is organized as follows. In Section 2, we derive the domination result of the Haff estimators over the usual estimators. We study numerically the behavior of the proposed estimators in Section 3.

2. Main result

Improving the class of the natural estimators in (3) relies on improving the optimal estimator among this class, that is, the one which minimizes the loss function (2).

Proposition 1 (Tsukuma [13]). *Under the Stein type loss function* (2), *the optimal estimator among the class* (3) *is given by*

$$\widehat{\boldsymbol{\Sigma}}_{a_o} = a_o\boldsymbol{S}, \quad \textit{where} \quad a_o = \frac{1}{m} \quad \textit{and} \quad m = \max(n, r). \tag{4}$$

As mentioned in Section 1, we consider the class of orthogonally invariant estimators as alternative estimators. Let $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^\top$ be the eigenvalue decomposition of $\boldsymbol{S}$ where $\boldsymbol{H}$ is a $p \times q$ semi–orthogonal matrix of eigenvectors and $\boldsymbol{L} = \operatorname{diag}(l_1, \dots, l_q)$, with $l_1 >, \dots, > l_q$, is the diagonal matrix of the q positive corresponding eigenvalues (see Kubokawa and Srivastava [7] for more details). The class of orthogonally invariant estimators is of the form

$$\widehat{\boldsymbol{\Sigma}}_\Psi = a_o \left(\boldsymbol{S} + \boldsymbol{H}\boldsymbol{L}\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{H}^\top\right) \tag{5}$$

with $\boldsymbol{\Psi}(\boldsymbol{L}) = \operatorname{diag}(\psi_1(\boldsymbol{L}), \dots, \psi_q(\boldsymbol{L}))$, where $\psi_i(\boldsymbol{L})$ $(i = 1, \dots, q)$ is a differentiable function of $\boldsymbol{L}$. More precisely, we consider an extension of the empirical Bayes Haff type estimators in [5], to the context of the model (1), defined as

$$\widehat{\boldsymbol{\Sigma}}_\alpha = a_o\left(\boldsymbol{S} + \boldsymbol{H}\boldsymbol{L}\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{H}^\top\right) \quad \text{where} \quad \boldsymbol{\Psi}(\boldsymbol{L}) = b\frac{\boldsymbol{L}^{-\alpha}}{\operatorname{tr}(\boldsymbol{L}^{-\alpha})}, \quad \text{with} \quad \alpha \ge 1 \quad \text{and} \quad b > 0, \tag{6}$$

where a_o is given in (4). Note that, for $\alpha = 1$, this is the estimator considered by Konno [6], who deals with the cases (i) and (ii) under a quadratic loss. An extension to elliptical setting was considered recently by Haddouche et al. [4]. Note also that Tsukuma [13] considered also empirical Bayes estimators but not the one in (6). We give in the following theorem our main result.

Theorem 2. *Let $\boldsymbol{\Psi}(\boldsymbol{L}) = \operatorname{diag}(\psi_1(\boldsymbol{L}), \dots, \psi_q(\boldsymbol{L}))$ where, for any $i = 1, \dots, q$, $\psi_i(\boldsymbol{L}) = b l_i^\alpha / \operatorname{tr}(\boldsymbol{L}^{-\boldsymbol{\alpha}})$ with $\alpha \ge 1$ and $b > 0$, be a differentiable function of $\boldsymbol{L}$. The Haff type estimators in* (6) *dominates the optimal estimator in* (4)*, under the loss function* (2)*, as soon as*

$$0 < b \le b_o = \frac{2\left(q-1\right)}{m-q+1}.$$

Proof. The proof consists to showing that the risk difference between the Haff type estimators in (6) and the optimal estimator in (4), namely,

$$\Delta_{(\alpha, a_o)} = R\left(\widehat{\boldsymbol{\Sigma}}_\alpha, \boldsymbol{\Sigma}\right) - R\left(\widehat{\boldsymbol{\Sigma}}_{a_o}, \boldsymbol{\Sigma}\right), \tag{7}$$

is non–positive. Note that $\widehat{\boldsymbol{\Sigma}}_\alpha$ can be written as

$$\widehat{\boldsymbol{\Sigma}}_\alpha = a_o\, \boldsymbol{H}\boldsymbol{L}\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{H}^\top \quad \text{with} \quad \boldsymbol{\Phi}(\boldsymbol{L}) = \boldsymbol{I}_q + \boldsymbol{\Psi}(\boldsymbol{L}).$$

The risk of these estimators under the Stein type loss function (2) is given by

$$R\left(\widehat{\boldsymbol{\Sigma}}_\alpha, \boldsymbol{\Sigma}\right) = E\left(\operatorname{tr}\left(\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}}_\alpha\right)\right) - E\left(\ln\left|\Lambda\left(\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}}_\alpha\right)\right|\right) - q. \tag{8}$$

First, dealing with $E(\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_\alpha\boldsymbol{\Sigma}^+))$, we apply in Tsukuma [13, Lemma A.2] in order to get rid of the unknown parameter $\boldsymbol{\Sigma}^+$. It follows that,

$$E\left(\operatorname{tr}\left(\boldsymbol{\Sigma}^+\widehat{\boldsymbol{\Sigma}}_\alpha\right)\right) = a_o E\left(\sum_{i=1}^q \left\{\left(m-q+1\right)\varphi_i + 2 l_i \frac{\partial \varphi_i}{\partial l_i} + 2\sum_{j>i}^q \frac{l_i\varphi_i - l_j\varphi_j}{l_i - l_j}\right\}\right) \tag{9}$$

where, for any $i = 1, \dots, q$,

$$\phi_i = 1 + b\,\frac{l_i^{-\alpha}}{\operatorname{tr}(L^{-\alpha})}, \quad \frac{\partial\varphi_i}{\partial l_i} = b\,\alpha\,\frac{1 - \operatorname{tr}(\boldsymbol{L}^{-\alpha})\; l_i^\alpha}{\operatorname{tr}^2(\boldsymbol{L}^{-\alpha})\; l_i^{1+2\alpha}}$$

and

$$\sum_{j>i}^q \frac{l_i\varphi_i - l_j\varphi_j}{l_i - l_j} = \sum_{j>i}^q \left(1 + \frac{b}{\operatorname{tr}(\boldsymbol{L}^{-\alpha})}\left\{\frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j}\right\}\right).$$

Using the fact, for any $j > i$, $l_j > l_i$, it can be shown that

$$\sum_{j>i}^q \frac{l_i\varphi_i - l_j\varphi_j}{l_i - l_i} \le \left(q - i\right). \tag{10}$$

Therefore, using (10), we obtain

$$
\begin{aligned}
E\left(\operatorname{tr}\left(\boldsymbol{\Sigma}^{+}\widehat{\boldsymbol{\Sigma}}_{\alpha}\right)\right) &\leq a_o\, E\left(\sum_{i=1}^{q}\left\{\left(m-q+1\right)\left(1+b\frac{l_i^{-\alpha}}{\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)}\right)+2\,b\,\alpha\,\frac{1-\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)\,l_i^{\alpha}}{\operatorname{tr}^2\left(\boldsymbol{L}^{-\alpha}\right)\,l_i^{2\alpha}}+2\left(q-i\right)\right\}\right) \\
&= a_o\, m\, q + a_o\, b\, E\left(\sum_{i=1}^{q}\left\{\left(m-q+1\right)\frac{l_i^{-\alpha}}{\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)}+2\,\alpha\,\frac{1-\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)\,l_i^{\alpha}}{\operatorname{tr}^2\left(\boldsymbol{L}^{-\alpha}\right)\,l_i^{2\alpha}}\right\}\right) \\
&= a_o\left(m\,q+b\left(m-q+1\right)\right)+2\,\alpha\, E\left(\frac{\operatorname{tr}\left(\boldsymbol{L}^{-2\alpha}\right)}{\operatorname{tr}^2\left(\boldsymbol{L}^{-\alpha}\right)}-1\right).
\end{aligned}
$$

Now, from the submultiplicativity of the trace for semi–definite positive matrices, we have $\operatorname{tr}(\boldsymbol{L}^{-2\alpha}) \leq \operatorname{tr}^2(\boldsymbol{L}^{-\alpha})$. Then, an upper bound for (9) is given by

$$
E\left(\operatorname{tr}\left(\boldsymbol{\Sigma}^{+}\widehat{\boldsymbol{\Sigma}}_{\alpha}\right)\right) \leq a_o\left(m\,q+b\left(m-q+1\right)\right). \tag{11}
$$

Secondly, dealing with $E(\ln|\Lambda(\widehat{\boldsymbol{\Sigma}}_{\alpha}\boldsymbol{\Sigma}^{+})|)$ in (8), it can be shown that

$$
\begin{aligned}
\Lambda\left(\boldsymbol{\Sigma}^{+}\widehat{\boldsymbol{\Sigma}}_{\alpha}\right) &= a_o\,\Lambda\left(\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{H}^{\top}\right) \\
&= a_o\,\Lambda\left(\boldsymbol{L}^{1/2}\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{\Phi}(\boldsymbol{L})\right).
\end{aligned}
$$

Note that $\boldsymbol{L}^{1/2}\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}^{1/2}$ and $\boldsymbol{\Phi}(\boldsymbol{L})$ are full rank $q\times q$ matrices. It follows that

$$
\left|\Lambda\left(\widehat{\boldsymbol{\Sigma}}_{\alpha}\boldsymbol{\Sigma}^{+}\right)\right| = a_o^{q}\left|\boldsymbol{L}^{1/2}\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{\Phi}(\boldsymbol{L})\right|.
$$

Therefore

$$
E\left(\ln\left|\Lambda\left(\widehat{\boldsymbol{\Sigma}}_{\alpha}\boldsymbol{\Sigma}^{+}\right)\right|\right) = q\,\ln(a_o)+E\left(\ln\left|\boldsymbol{L}^{1/2}\boldsymbol{H}^{T}\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}^{1/2}\right|\right)+E\left(\ln|\boldsymbol{\Phi}(\boldsymbol{L})|\right). \tag{12}
$$

Using the fact that $\ln(1+x)\geq 2\,x/(2+x)$, for any positive constant x, we have

$$
\begin{aligned}
\ln|\boldsymbol{\Phi}(\boldsymbol{L})| &= \ln\left|\boldsymbol{I}_q+\frac{b\,\boldsymbol{L}^{-\alpha}}{\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)}\right| \\
&= \sum_{i=1}^{q}\ln\left(1+\frac{b\,l_i^{-\alpha}}{\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)}\right) \\
&\geq \sum_{i=1}^{q}\frac{2\,b\,l_i^{-\alpha}/\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)}{2+b\,l_i^{-\alpha}/\operatorname{tr}\left(\boldsymbol{L}^{-\alpha}\right)}.
\end{aligned}
$$

Then

$$
\ln|\boldsymbol{\Phi}(\boldsymbol{L})| \geq \frac{2b}{2+b}, \tag{13}
$$

since, for any $i=1,\ldots,q$, $l_i^{-\alpha}\leq\operatorname{tr}(\boldsymbol{L}^{-\alpha})$. Consequently, thanks to (13), a lower bound for (12) is given by

$$
E\left(\ln\left|\Lambda\left(\widehat{\boldsymbol{\Sigma}}_{\alpha}\boldsymbol{\Sigma}^{+}\right)\right|\right) \geq q\,\ln(a_o)+E\left(\ln\left|\boldsymbol{L}^{1/2}\boldsymbol{H}^{T}\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}^{1/2}\right|\right)+\frac{2\,b}{2+b}. \tag{14}
$$

Now, relying on the proof of [13, Proposition 2.1] in Tsukuma, it can be shown that

$$
R\left(\widehat{\boldsymbol{\Sigma}}_{a_o},\boldsymbol{\Sigma}\right) = -q\ln(a_o)-E\left(\ln\left|\boldsymbol{L}^{1/2}\boldsymbol{H}^{T}\boldsymbol{\Sigma}^{+}\boldsymbol{H}\boldsymbol{L}^{1/2}\right|\right). \tag{15}
$$

Finally, combining (11), (14) and (15), an upper bound for the risk difference (7) is given by

$$
\begin{aligned}
\Delta_{(\alpha,a_o)} &\leq a_o\left(m-q+1\right)b-\frac{2\,b}{2+b} \\
&= \frac{b}{b+2}\left(\frac{\left(m-q-1\right)\left(b+2\right)}{m}-2\right),
\end{aligned}
$$

since $a_o=1/m$, which is non-positive as soon as

$$
0<b\leq b_o=\frac{2\left(q-1\right)}{m-q+1}. \qquad \square
$$

3. Numerical study

We study here numerically the performance of the Haff type estimators in (6). As shown numerically by Fourdrinier et al. [3] that the best constant b is b_o. So that, we consider in this section the following Haff type estimators

$$\widehat{\boldsymbol{\Sigma}}_\alpha = a_o\left(\boldsymbol{S} + \boldsymbol{H}\boldsymbol{L}\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{H}^\top\right) \quad \text{where} \quad \boldsymbol{\Psi}(\boldsymbol{L}) = b_o\frac{\boldsymbol{L}^{-\alpha}}{\operatorname{tr}(\boldsymbol{L}^{-\alpha})},$$

$$\text{with} \quad \alpha \geq 1 \quad \text{and} \quad b_o = \frac{2\,(q-1)}{m-q+1}. \qquad (16)$$

As for the population covariance matrix $\boldsymbol{\Sigma}$, we consider the following structures: (i) the identity matrix $\boldsymbol{I}_p$ and (ii) an autoregressive structure with coefficient 0.9. We set their $p-r$ smallest eigenvalues to zero in order to construct matrices of rank $r \leq p$.

To assess the performance of the proposed estimators, we compute the Percentage Reduction In Average Loss (PRIAL), for some values of p, n, r and α, defined as

$$\operatorname{PRIAL}\left(\widehat{\boldsymbol{\Sigma}}_\alpha\right) = \frac{R\left(\widehat{\boldsymbol{\Sigma}}_{a_o},\boldsymbol{\Sigma}\right) - R\left(\widehat{\boldsymbol{\Sigma}}_\alpha,\boldsymbol{\Sigma}\right)}{R\left(\widehat{\boldsymbol{\Sigma}}_{a_o},\boldsymbol{\Sigma}\right)} \times 100,$$

where $\widehat{\boldsymbol{\Sigma}}_{a_o}$ and $\widehat{\boldsymbol{\Sigma}}_\alpha$ are respectively defined in (4) and (3).

Table 1. Effect of α on PRIAL's for the structures(i) and (ii) of $\boldsymbol{\Sigma}$.

$\boldsymbol{\Sigma}$	(p,n)	r	$\widehat{\boldsymbol{\Sigma}}_1$	$\widehat{\boldsymbol{\Sigma}}_2$	$\widehat{\boldsymbol{\Sigma}}_3$	$\widehat{\boldsymbol{\Sigma}}_4$	$\widehat{\boldsymbol{\Sigma}}_5$
(i)	(30,50)	10	6.85	12.45	15.53	16.95	17.53
		20	9.20	13.91	14.88	14.68	12.47
		30	11.81	14.33	13.41	12.43	11.71
	(50,30)	20	18.31	19.65	17.75	16.44	15.63
		40	17.12	16.33	14.07	12.78	12.02
		50	11.80	14.23	13.29	12.30	11.59
	(150,30)	20	18.17	19.69	17.87	16.56	15.71
		40	17.08	16.31	14.07	12.76	11.99
		60	8.88	12.27	12.33	11.75	11.19
		150	2.83	5.09	6.50	7.25	7.61
(ii)	(30,50)	10	6.06	8.70	9.62	9.89	9.92
		20	8.81	11.66	12.12	11.93	11.61
		30	11.46	13.15	12.35	11.48	10.82
	(50,30)	20	17.18	17.45	15.85	14.76	14.07
		40	16.34	15.18	13.19	12.00	11.28
		50	11.33	12.82	12.02	11.19	10.57
	(150,30)	20	17.30	18.00	16.38	15.19	14.42
		40	16.27	15.04	13.04	11.84	11.13
		60	8.48	10.43	10.29	9.81	9.35
		150	2.73	4.03	4.61	4.86	4.95

Table 1 shows that the proposed estimators improve over $\widehat{\boldsymbol{\Sigma}}_{a_o}$ for any possible ordering of p, n and r. Compared to other cases, the Haff type estimators $\widehat{\boldsymbol{\Sigma}}_\alpha$ (for $\alpha = 1,\ldots,5$) have better performances in the case where $p > n > r$, with PRIAL's higher than 14.07% for both structures (i) and (ii) of $\boldsymbol{\Sigma}$. We report that the optimal value of α, which maximizes the PRIAL's, depends on p, n and r.

We then compare the Prial's of the Haff type estimators in and the well known James–Stein type estimator $\widehat{\boldsymbol{\Sigma}}_{JS}$ considered by Tsukuma [13], given by

$$\widehat{\boldsymbol{\Sigma}}_{JS} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{D}_q^{JS}\boldsymbol{H}^\top,$$

where $\boldsymbol{D}_q^{JS} = \mathrm{diag}(d_1^{JS}, \dots, d_q^{JS})$ with, for $i = 1, \dots, q$, $d_i^{JS} = (n + r - 2i + 1)^{-1}$.

Table 2. PRIAL's of the Haff type estimators and the Stein type estimator for the structures (ii) of $\boldsymbol{\Sigma}$.

(p,n)	r	$\widehat{\boldsymbol{\Sigma}}_1$	$\widehat{\boldsymbol{\Sigma}}_2$	$\widehat{\boldsymbol{\Sigma}}_3$	$\widehat{\boldsymbol{\Sigma}}_4$	$\widehat{\boldsymbol{\Sigma}}_{JS}$
$(30,40)$	7	9.06	12.34	13.36	13.68	9.87
	14	9.60	13.09	13.78	13.74	20.45
$(40,30)$	7	11.87	15.57	16.41	16.55	16.04
	14	13.91	16.41	16.13	15.55	26.35
$(20,100)$	5	4.04	5.61	6.14	6.29	4.68
	15	3.67	5.45	6.38	7.12	11.80

Table 2 shows that in some cases the PRIALs of the Haff-type estimators are higher than those of the Stein-type estimator, while in other cases the Stein-type estimator has a better performance. Consequently, their risk functions intersect and thus, in our opinion, a theoretical dominance results between these two estimators cannot be derived.

References

[1] D. Chételat, M. T. Wells, "Improved second order estimation in the singular multivariate normal model", *J. Multivariate Anal.* **147** (2016), p. 1-19.

[2] D. K. Dey, C. Srinivasan, "Estimation of a covariance matrix under Stein's loss", *Ann. Stat.* **13** (1985), p. 1581-1591.

[3] D. Fourdrinier, A. M. Haddouche, F. Mezoued, "Covariance matrix estimation under data-based loss", *Stat. Probab. Lett.* **177** (2021), article no. 109160.

[4] A. M. Haddouche, D. Fourdrinier, F. Mezoued, "Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions", *J. Multivariate Anal.* **181** (2021), article no. 104680.

[5] L. R. Haff, "Empirical Bayes estimation of the multivariate normal covariance matrix", *Ann. Stat.* **8** (1980), no. 3, p. 586-597.

[6] Y. Konno, "Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss", *J. Multivariate Anal.* **100** (2009), no. 10, p. 2237-2253.

[7] T. Kubokawa, M. S. Srivastava, "Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data", *J. Multivariate Anal.* **99** (2008), no. 9, p. 1906-1928.

[8] O. Ledoit, M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices", *J. Multivariate Anal.* **88** (2004), no. 2, p. 365-411.

[9] M. S. Srivastava, "Singular Wishart and multivariate Beta distributions", *Ann. Stat.* **31** (2003), no. 5, p. 1537-1560.

[10] C. M. Stein, "Estimation of a covariance matrix", in *Rietz Lecture, 39th Annual Meeting of the IMS, Atlanta, GA, 1975*, 1975.

[11] ———, "Lectures on the theory of estimation of many parameters", *J. Sov. Math.* **34** (1986), no. 1, p. 1373-1403.

[12] A. Takemura, "An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population", *Tsukuba J. Math.* **8** (1984), p. 367-376.

[13] H. Tsukuma, "Estimation of a high-dimensional covariance matrix with the Stein Loss", *J. Multivariate Anal.* **148** (2016), p. 1-17.