Statistics and Probability Letters 177 (2021) 109160

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Covariance matrix estimation under data-based loss

Dominique Fourdrinier[a], Anis M. Haddouche[b,*], Fatiha Mezoued[c]

[a] *Université de Normandie, Normandie Univ, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, avenue de l'Université, BP 12, Saint-Étienne-du-Rouvray, 76801, France*
[b] *INSA Rouen, Normandie Univ, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS and LMI, avenue de l'Université, BP 8, Saint-Étienne-du-Rouvray, 76801, France*
[c] *École Nationale Supérieure de Statistique et d'Économie Appliquée (ENSSEA), LAMOPS, Tipaza, Algeria*

ARTICLE INFO

Article history:
Received 22 December 2020
Received in revised form 12 March 2021
Accepted 14 May 2021
Available online 26 May 2021

MSC:
62H12
62F10
62C99

Keywords:
Data-based loss
Elliptically symmetric distributions
High-dimensional statistics
Orthogonally invariant estimators
Stein-Haff type identities

ABSTRACT

We consider here the problem of estimating the $p \times p$ scale matrix Σ of a multivariate linear regression model when the distribution of the observed matrix belongs to a large class of elliptically symmetric distributions. Any estimator $\hat{\Sigma}$ of Σ is assessed through the data-based loss $\operatorname{tr}\left(S^{+}\Sigma\,(\Sigma^{-1}\hat{\Sigma}-I_p)^2\right)$ where S is the sample covariance matrix and S^{+} is its Moore–Penrose inverse.

1. Introduction

Let consider the multivariate linear regression model, with p responses and n observations,

$$Y = X\,\beta + \mathcal{E}\,, \tag{1}$$

where Y is an $n \times p$ matrix, X is an $n \times q$ matrix of known constants of rank $q \leq n$ and β is a $q \times p$ matrix of unknown parameters. We assume that the $n \times p$ noise matrix $\mathcal{E}$ has an elliptically symmetric distribution with density, with respect to the Lebesgue measure in $\mathbb{R}^{pn}$, of the form

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\left(\operatorname{tr}(\,\varepsilon\,\Sigma^{-1}\varepsilon^{\top})\right), \tag{2}$$

where Σ is a $p \times p$ unknown positive definite matrix and $f(\cdot)$ is a non-negative unknown function.

The model (1) has been considered by various authors such as Kubokawa and Srivastava (1999, 2001), who estimated Σ and β respectively in the context (2), and Tsukuma and Kubokawa (2016), who estimated Σ in the Gaussian setting. A common alternative representation of this model is $Y = M + \mathcal{E}$, where $\mathcal{E}$ is as above and M is in the column space of X, has been also considered in the literature. See for instance Candès et al. (2013) and Canu and Fourdrinier (2017).

* Corresponding author.
E-mail addresses: dominique.fourdrinier@univ-rouen.fr (D. Fourdrinier), mohamed.haddouche@insa-rouen.fr (A.M. Haddouche), famezoued@yahoo.fr, mezoued.fatiha@enssea.net (F. Mezoued).

https://doi.org/10.1016/j.spl.2021.109160
0167-7152/

Although the matrix of regression coefficients β is also unknown, we are interested in estimating the scale matrix Σ. We address this problem under a decision-theoretic framework through a canonical form of the model (1), which allows to use a sufficient statistic $S = U^\top U$ for Σ, where U is an $(n-q) \times p$ matrix (see Section 2 for more details). In this context, the natural estimators of Σ are of the form

$$\hat{\Sigma}_a = a\,S \quad \text{where} \quad a > 0. \tag{3}$$

As pointed out by James and Stein (1961), the estimators of the form (3) perform poorly in the Gaussian setting. In fact, larger (smaller) eigenvalues of Σ are overestimated (underestimated) by those estimators. Thus we may expect to improve these estimators by shrinking the eigenvalues of S, which gives rise to the class of orthogonally invariant estimators (see Takemura, 1984). Since the seminal work of James and Stein (1961), this problem has been largely considered in the Gaussian setting. See, for instance, Tsukuma (2016) and Chételat and Wells (2016). However, the elliptical setting has been considered by a few authors such as Kubokawa and Srivastava (1999) and Haddouche et al. (2021).

In this paper, the performance of any estimator $\hat{\Sigma}$ of Σ is assessed through the data-based loss

$$L_S(\hat{\Sigma}, \Sigma) = \operatorname{tr}\Big(S^+ \Sigma \left(\Sigma^{-1}\hat{\Sigma} - I_p\right)^2\Big) \tag{4}$$

and its associated risk

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}\big[L_S(\hat{\Sigma}, \Sigma)\big]\,, \tag{5}$$

where $E_{\theta,\Sigma}$ denotes the expectation with respect to the density specified below in (9) and where S^+ is the Moore–Penrose inverse of S. Note that, when $p > n-q$, S is non-invertible and, when $p \leq n-q$, S is invertible so that S^+ coincides with the regular inverse S^{-1}. This type of loss is called data-based loss in so far as it contains a part of the observation U through $S = U^\top U$. The notion of data-based loss was introduced by Efron and Morris (1976) when estimating a location parameter. Likewise, Fourdrinier and Strawderman (2015) showed the interest of considering such a data-based loss with respect to the usual quadratic loss (see (6) below). Also, the data-based loss (4) was considered, in a Gaussian setting, by Tsukuma and Kubokawa (2015) who were motivated by the difficulty to handle the standard quadratic loss

$$L(\hat{\Sigma}, \Sigma) = \operatorname{tr}\Big(\Sigma^{-1}\hat{\Sigma} - I_p\Big)^2\,. \tag{6}$$

See Tsukuma (2016) for more details. Thus the loss in (4) is a data-based variant of (6), through which we aim to improve on the estimators $\hat{\Sigma}_a$ in (3) using alternative estimators, focusing on orthogonally invariant estimators. Note that most improvement results, in the Gaussian case, were derived thanks to Stein–Haff type identities. Here, we specifically use the Stein–Haff type identity given by Haddouche et al. (2021), in the elliptical case, to establish our dominance result, which is well adapted to our unified approach to the cases S invertible and S non-invertible.

The rest of this paper is structured as follows. In Section 2, we give conditions for improving the proposed estimators over the usual estimators. In Section 3, we assess the quality of the proposed estimators through a simulation study in the context of the t-distribution. We also compare numerically our results with those of Konno (2009) in the Gaussian setting. We give, in an Appendix (given as a supplementary material), all the proofs of our findings.

2. Main result

Although we are interested in estimating the scale matrix Σ, recall that β is a $q \times p$ matrix of unknown parameters. Note that, since X has full column rank, the least squares estimator of β is $\hat{\beta} = (X^\top X)^{-1} X^\top Y$; this is the maximum likelihood estimator in the Gaussian setting. Natural estimators of the scale matrix Σ are based on the residual sum of squares given by

$$S = Y^\top (I_n - P_X)\, Y, \tag{7}$$

where $P_X = X(X^\top X)^{-1} X^\top$ is the orthogonal projector onto the subspace spanned by the columns of X.

Following the lines of Tsukuma and Kubokawa (2020), we derive the canonical form of the model (1) which allows a suitable treatment of the estimation of Σ. Let $X = Q_1 T^\top$ be the QR decomposition of X where Q_1 is a $n \times q$ semi-orthogonal matrix and T a $q \times q$ lower triangular matrix with positive diagonal elements. Setting $m = n - q$, there exists a $n \times m$ semi-orthogonal matrix Q_2 which completes Q_1 such that $Q = (Q_1 Q_2)$ is an $n \times n$ orthogonal matrix. Then, since $Q_2^\top X \beta = Q_2^\top Q_1 T^\top \beta = 0$, we have

$$Q^\top Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \left(Q_1^\top\right) X \beta + Q^\top \mathcal{E} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E}\,, \tag{8}$$

where $Q_1^\top X \beta = \theta$ and where Z and U are, respectively, $q \times p$ and $m \times p$ matrices. As $X = Q_1 L^\top$, the projection matrix P_X satisfies $P_X = Q_1 L^\top (L^\top L)^{-1} L\, Q_1^\top = Q_1 Q_1^\top$ so that $I_n - P_X = Q_2 Q_2^\top$. It follows that (7) becomes $S = Y^\top Q_2 Q_2^\top Y = U^\top U$, according to (8), which is a sufficient statistic for Σ.

The orthogonal matrix Q provides a linear reduction from n to q observations within each of the p responses. In addition, according to (2), the density of $Q^\top \mathcal{E}$ is the same as that of $\mathcal{E}$, and hence, $(Z^\top U^\top)^\top$ has an elliptically symmetric distribution about the matrix $(\theta^\top 0^\top)^\top$ with density

$$(z,u) \mapsto |\Sigma|^{-n/2} f\left(\operatorname{tr}(z-\theta)\,\Sigma^{-1}(z-\theta)^\top + \operatorname{tr} u\,\Sigma^{-1} u^\top\right), \tag{9}$$

where θ and Σ are unknown. In this sense, the model (8) is the canonical form of the multivariate linear regression model (1). Note that the marginal distribution of $U = Q_2^\top Y$ is elliptically symmetric about 0 with covariance matrix proportional to $I_m \otimes \Sigma$ (see Fang and Zhang, 1990). This implies that $S = U^\top U$ has a generalized Wishart distribution (see Díaz-Gacía and Gutiérrez-Jámez, 2011), which coincides with the standard (singular or non-singular) Wishart distribution in the Gaussian setting (see Srivastava, 2003).

As mentioned in Section 1, we propose alternative estimators; they are of the form

$$\hat{\Sigma}_J = a(S+J), \tag{10}$$

where $J = J(Z,S)$ appears as a correction matrix. The improvement over the class of estimators $\hat{\Sigma}_a$ can be done by improving over the best estimator $\hat{\Sigma}_{a_o} = a_o S$ within this class, namely, the estimator which minimizes the risk (5). It is proved in the Appendix that

$$\hat{\Sigma}_{a_o} = a_o S, \quad \text{with} \quad a_o = \frac{1}{K^* v} \quad \text{and} \quad v = \max\{p,m\}, \tag{11}$$

where K^* is the normalizing constant (assumed to be finite) of the density defined by

$$(z,u) \mapsto \frac{1}{K^*} |\Sigma|^{-n/2} F^*\left(\operatorname{tr}(z-\theta)\,\Sigma^{-1}(z-\theta)^\top + \operatorname{tr} u\,\Sigma^{-1} u^\top\right), \tag{12}$$

where, for any $t \geq 0$, $F^*(t) = \frac{1}{2}\int_t^\infty f(v)\,dv$. Note that under the loss (6) the optimal constant is $1/K^*(p+m+1)$. Of course, this risk optimality has sense only if the risk of $\hat{\Sigma}_{a_o}$ is finite. As shown in Haddouche (2019), this is the case as soon as $E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}S\right)\right] < \infty$ and $E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma\, S^+\right)\right] < \infty$, which can be reduced to $E_{\theta,\Sigma}\left[\operatorname{tr}\left(S\right)\right] < \infty$ and $E_{\theta,\Sigma}\left[\operatorname{tr}\left(S^+\right)\right] < \infty$ by submultiplicativity of the trace for semi-definite positive matrices. In order to give a unified dominance result of $\hat{\Sigma}_J$ over $\hat{\Sigma}_{a_o}$ for the two cases where S is non-invertible and where S is invertible, we consider, as a correction matrix in (10), the projection of a matrix function $G(Z,S) = G$ on the subspace spanned by the columns of SS^+, namely,

$$J = SS^+G. \tag{13}$$

In addition to the risk finiteness conditions of $\hat{\Sigma}_{a_o}$, it can be shown that the risk of $\hat{\Sigma}_J$ is finite as soon as the expectations $E_{\theta,\Sigma}\left[\|\,\Sigma^{-1}SS^+G\|_F^2\right]$ and $E_{\theta,\Sigma}\left[\|S^+G\|_F^2\right]$ are finite, where $\|\cdot\|_F$ denotes the Frobenius norm. Note that $E_{\theta,\Sigma}\left[\|\,\Sigma^{-1}SS^+G\|_F^2\right] < \infty$ reduces to $E_{\theta,\Sigma}\left[\|SS^+G\|_F^2\right] < \infty$ by submultiplicativity of the Frobenius norm. Under these conditions, the risk difference between $\hat{\Sigma}_J$ and $\hat{\Sigma}_{a_o}$ is

$$\Delta(G) = a_o^2 E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}\, SS^+\, G\{I_p + S^+G + SS^+\}\right)\right] - 2\,a_o\, E_{\theta,\Sigma}\left[\operatorname{tr}\left(S^+\, G\right)\right]. \tag{14}$$

Noticing that the first integrand term in (14) depends on the unknown parameter Σ^{-1}, our approach consists in replacing this integrand term by a random matrix $\delta(G)$, which does not depend on Σ^{-1}, such that $E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}\, SS^+\, G\{I_p + S^+G + SS^+\}\right)\right] = E^*_{\theta,\Sigma}\left[\delta(G)\right]$, where $E^*_{\theta,\Sigma}$ denotes the expectation with respect to the density (12). To this end, we rely on the following Stein–Haff type identity, which is based on the notion of weak differentiability naturally involved in Stein's lemma (see Fourdrinier et al., 2018 for more details).

Lemma 1 (*Haddouche et al., 2021*)**.** *Let $G(z,s)$ be a $p \times p$ matrix function such that, for any fixed z, $G(z,s)$ is weakly differentiable with respect to s. Assume that $E_{\theta,\Sigma}\left[|\operatorname{tr}(\Sigma^{-1} S\, S^+\, G)|\right] < \infty$. Then we have*

$$E_{\theta,\Sigma}\left[\operatorname{tr}\left(\,\Sigma^{-1}\, SS^+\, G\right)\right] = K^*\, E^*_{\theta,\Sigma}\left[\operatorname{tr}\left(2\, SS^+\, \mathcal{D}_s\{SS^+G\}^\top + (m-r-1)\, S^+\, G\right)\right], \tag{15}$$

where $r = \min\{p,m\}$ and $\mathcal{D}_s\{\cdot\}$ 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 $i \neq j$.

Note that the existence of the expectations in (15) is implied by the above risk finiteness conditions. Applying Lemma 1 to the term depending on Σ^{-1} on the right-hand side of (14) gives

$$\begin{aligned} \Delta(G) = a_o^2\, K^*\, E^*_{\theta,\Sigma}\Big[&(m-r-1)\operatorname{tr}\left(S^+G + (S^+G)^2 + S^+GSS^+\right) \\ &+ 2\operatorname{tr}\left(SS^+\, \mathcal{D}_s\{SS^+G + SS^+GS^+G + SS^+\, G\, SS^+\}^\top\right)\Big] - 2\, a_o\, E_{\theta,\Sigma}\left[\operatorname{tr}\left(S^+\, G\right)\right]. \end{aligned} \tag{16}$$

It is worth noticing that the risk difference in (16) depends on the $E_{\theta,\Sigma}$ and $E^*_{\theta,\Sigma}$ expectations (which coincide in the Gaussian setting since $F^* = f$). Thus, in order to derive a dominance result, we need to compare these two expectations. A possible approach consists in restricting us to the subclass of densities verifying $c \leq F^*(t)/f(t) \leq b$, for some positive constants c and b (see Berger, 1975 for the class where $c \leq F^*(t)/f(t)$). Due to the complexity of the use of the quadratic loss in (6) (which necessitates a twice application of the Stein–Haff type identity (15)), this subclass was considered by Haddouche et al. (2021). Here, thanks to the data-based loss (4), we are able to avoid such a restriction, and hence, to deal with a larger class of elliptically symmetric distributions in (9) (subject to the moment conditions induced by the above finiteness conditions).

Following the suggestion to shrink the eigenvalues of S mentioned in Section 1, we consider as a correction matrix a matrix SS^+G with G orthogonally invariant in the following sense. Let $S = H\,L\,H^\top$ be the eigenvalue decomposition of S where H is a $p \times r$ semi-orthogonal matrix of eigenvectors and $L = \mathrm{diag}(l_1, \ldots, l_r)$, with $l_1 >, \ldots, > l_r$, is the diagonal matrix of the r positive corresponding eigenvalues of S (see Kubokawa and Srivastava, 2008 for more details). Then set $G = H\,L\Psi(L)\,H^\top$, with $\Psi(L) = \mathrm{diag}(\psi_1(L), \ldots, \psi_r(L))$ where $\psi_i = \psi_i(L)$ $(i = 1, \ldots, r)$ is a differentiable function of L. Consequently, by semi-orthogonality of H, we have $SS^+H = H\,H^\top H = H$, so that the correction matrix in (13) is $J = SS^+G = G = H\,L\Psi(L)\,H^\top$ and $S^+G = H\,\Psi(L)\,H^\top$. Thus the alternative estimators that we consider are of the form

$$\hat{\Sigma}_\Psi = a_o\left(S + H\,L\,\Psi(L)\,H^\top\right) = a_o\,H\,L\left(I_r + \Psi(L)\right)H^\top, \tag{17}$$

which are usually called orthogonally invariant estimators (i.e. equivariant under orthogonal transformations). See for instance Takemura (1984).

Now, adapting the risk finiteness conditions mentioned above, we are in a position to give our dominance result of the alternative estimators in (17) over the optimal estimator in (11), under the data-based loss (4).

Theorem 1. *Assume that the expectations $E_{\theta,\Sigma}\big[\mathrm{tr}(S)\big]$, $E_{\theta,\Sigma}\big[\mathrm{tr}(S^+)\big]$, $E_{\theta,\Sigma}\big[\|HL\Psi(L)\,H^\top\|_F^2\big]$ and $E_{\theta,\Sigma}\big[\|H\Psi(L)H^\top\|_F^2\big]$ are finite. Let $\Psi(L) = \mathrm{diag}(\psi_1, \ldots, \psi_r)$ where $\psi_i = \psi_i(L)$ $(i = 1, \ldots, r)$ is differentiable function of L with $\mathrm{tr}\big(\Psi(L)\big) \geq \lambda$, for a fixed positive constant λ.*

Then an upper bound of the risk difference between $\hat{\Sigma}_\Psi$ and $\hat{\Sigma}_{a_o}$ under the loss (4) is given by

$$\Delta(\Psi(L)) \leq a_o^2\,K^*\,E^*_{\theta,\Sigma}\big[g(\Psi)\big],$$

where

$$g(\Psi) = \sum_{i=1}^{r}\Bigg\{2(v-r+1)\psi_i + (v-r+1)\psi_i^2 + 4l_i(1+\psi_i)\frac{\partial\psi_i}{\partial l_i} + \sum_{j\neq i}^{r}\frac{l_i(2\psi_i+\psi_i^2) - l_j(2\psi_j+\psi_i^2)}{l_i - l_j} - 2v\lambda\Bigg\}. \tag{18}$$

Also, $\hat{\Sigma}_\Psi$ in (17) improves over $\hat{\Sigma}_{a_o}$ in (11) as soon as $g(\Psi) \leq 0$.

The proof of Theorem 1 is given in the Appendix. Note that the fact that $E_{\theta,\Sigma}\,[\mathrm{tr}(S)] < \infty$ and $E_{\theta,\Sigma}\,\big[\mathrm{tr}(S^+)\big] < \infty$ is necessary for the estimators $\hat{\Sigma}_a$ to have finite risk. This is a condition on the family of distributions. Note also that the assumptions $E_{\theta,\Sigma}\big[\|HL\Psi(L)H^\top\|_F^2\big] < \infty$ and $E_{\theta,\Sigma}\big[\|H\Psi(L)H^\top\|_F^2\big] < \infty$ guarantee the finiteness of the risk of $\hat{\Sigma}_\Psi$.

Although the expectation $E^*_{\theta,\Sigma}$ is associated, through $F^*(\cdot)$ in (12), to the generating function $f(\cdot)$ in (2), the function $g(\Psi)$ does not depend on $f(\cdot)$, and hence, the improvement result in Theorem 1 is robust in that sense. Note that Theorem 1 is well adapted to deal with the James and Stein (1961) estimator where $\psi_i(L) = 1/(v + r - 2i + 1)$, for $i = 1, \ldots, r$, since $\mathrm{tr}\big(\Psi(L)\big) > \lambda = 1/(v + r - 1)$, and the Efron–Morris–Dey estimator, considered by Tsukuma and Kubokawa (2020), where $\psi_i(L) = 1/\big(1 + b\,l_i^\alpha/\mathrm{tr}(L^\alpha)\big)v$, for $i = 1, \ldots, r$ and for positive constants b and α, since $\mathrm{tr}\big(\Psi(L)\big) > \lambda = r\,/(b+1)\,v$.

In the following, as a general example, we consider a new class of estimators which is an extension of the Haff (1980) class, that is, estimators of the form

$$\hat{\Sigma}_{\alpha,b} = a_o\left(S + H\,L\,\Psi(L)\,H^\top\right) \text{ with, for } \alpha \geq 1 \text{ and } b > 0,\ \ \Psi(L) = b\,\frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}, \tag{19}$$

where a_o is given in (11). For $\alpha = 1$, this is the estimator considered by Konno (2009), who deals with the Gaussian case and the quadratic loss (6), while Tsukuma and Kubokawa (2020) used an extended Stein loss. An elliptical setting was also considered by Haddouche et al. (2021) under the quadratic loss (6).

It is proved in the Appendix that the finiteness risk conditions of $\hat{\Sigma}_{\alpha,b}$ is reduced to $E_{\theta,\Sigma}[\mathrm{tr}^2(S)] < \infty$. It is also proved that any estimator $\hat{\Sigma}_{\alpha,b}$ in (19) improves on the optimal estimator $\hat{\Sigma}_{a_o}$ in (11), under the data-based loss (4), as soon as

$$0 < b \leq \frac{2\,(r-1)}{v - r + 1}\,. \tag{20}$$

It is worth noting that Tsukuma and Kubokawa (2020) gave the double inequality in Condition (20) as an improvement condition although their loss was different.

3. Numerical study

Let the elliptical density in (2) be a variance mixture of normal distributions where the mixing variable has the inverse-gamma distribution with shape and inverse scale parameters both equal to $k/2$ for $k > 2$. Thus, for any $t \geq 0$, the generating function f in (2) corresponds to the t-distribution with k degrees of freedom (see the Appendix). In the following, we study numerically the performance of the alternative estimators in (19) expressed as

$$\hat{\Sigma}_{\alpha,b} = a_o \left(S + \frac{b}{\mathrm{tr}(L^{-\alpha})} H \, L^{1-\alpha} \, H^\top \right) \tag{21}$$

where

$$0 \leq b \leq b_0 = \frac{2\,(r-1)}{v-r+1}$$

and

$$\alpha \geq 1.$$

As mentioned above, Konno (2009) considered the case $\alpha = 1$, in the Gaussian setting and under the quadratic loss (6), for which its improvement condition is $0 \leq b \leq b_1 = 2\,(r-1)\,(v+r+1)/(v-r+1)\,(v-r+3)$. Note that, although $b_0 < b_1$, the improvement condition in (21) is valid fo any $\alpha \geq 1$ and all the class of elliptically symmetric distributions (9). However it was shown numerically by Haddouche et al. (2021) that b_1 is optimal in the Gaussian context.

We consider the following structures of Σ: (i) the identity matrix I_p and (ii) an autoregressive structure with coefficient 0.9 (i.e. a $p \times p$ matrix where the (i, j)th element is $0.9^{|i-j|}$). To assess how an alternative estimator $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$, we compute the PRIAL (see the Appendix) based on 1000 independent Monte Carlo replications for some couples (p, m).

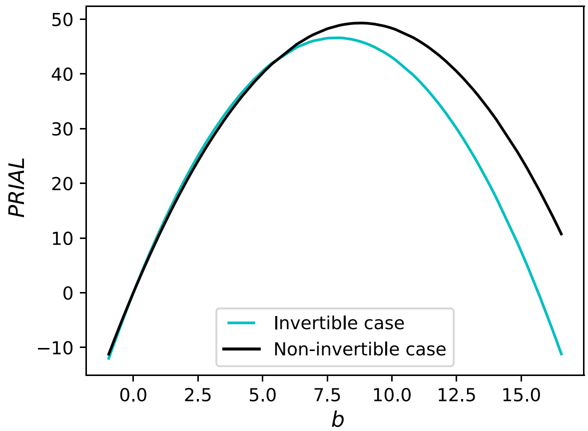

Fig. 1. Effect of b on the PRIAL of $\hat{\Sigma}_{\alpha,b}$, with $\alpha = 1$, under data-based loss in the Gaussian setting. The structure (i) of Σ is considered for $(p, m) = (10, 25)$ and $(p, m) = (25, 10)$.

In Fig. 1, we study the effect of the constant b in (21) on the prials in the non-invertible ($(p, m) = (25, 10)$) and the invertible ($(p, m) = (10, 25)$) cases. The Gaussian setting is investigated for the structure (i) of Σ. Note that, when $0 \leq b \leq b_0$, the best prial (around 7% in both invertible and non-invertible cases) is reported for $b = b_0 = 1.125$ (for $(v, r) = (25, 20)$). For this reason, in the following, we consider the estimators $\hat{\Sigma}_{\alpha,b_0}$ with $b_0 = 2\,(r-1)/(v-r+1)$. Note also that, for $b > b_0$, the estimators $\hat{\Sigma}_{\alpha,b}$ still improve over $\hat{\Sigma}_{a_o}$ and that the maximum value of the prial is around 50%. This shows that there exists a range of values of b, larger than the one that our theory provides, for which $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$.

In Fig. 2, we study the effect of α on the prials of the estimator $\hat{\Sigma}_{\alpha,b_0}$ over $\hat{\Sigma}_{a_o} = S/v$ when the sampling distribution is Gaussian ($K^* = 1$ in (11)), and over $\hat{\Sigma}_{a_o} = S(k-2)/vk$ when it is the t-distribution (K^* in (11) equals $(k-2)/k$ according to the Appendix) with k degrees of freedom. For the structure i of Σ, note that, for $\alpha \geq 6$, the prials stabilize at 12.5%, in the Gaussian case, and at 8.5%, in the Student's case. Similarly, the prials are better in the Gaussian setting for the structure (ii). In addition, it is interesting to observe that, when α is close to zero, the prials are small for the structure (i) and may be negative for the structure (ii).

In Fig. 3, under the Gaussian assumption, we provide the prials of $\hat{\Sigma}_{\alpha,b_0}$ with respect to $\hat{\Sigma}_{a_o} = S/v$ under the data-based loss (4) and the prials of $\hat{\Sigma}_{\alpha,b_1}$ with respect to $\hat{\Sigma}_{a_o} = S/(v+r+1)$ under the quadratic loss (6). For the structures (i) and (ii), the prials are better under the data-based loss. For the structure (i) with $\alpha = 1$ (which coincide with the Konno's estimator), we observe a prial equal to 1.73% which is similar to that of Konno (2009). Note that, under the data-based loss the prial is much better since it equals 13.42%. We observe similar behaviors for the structure (ii) than for the structure (i), but with lower prials.

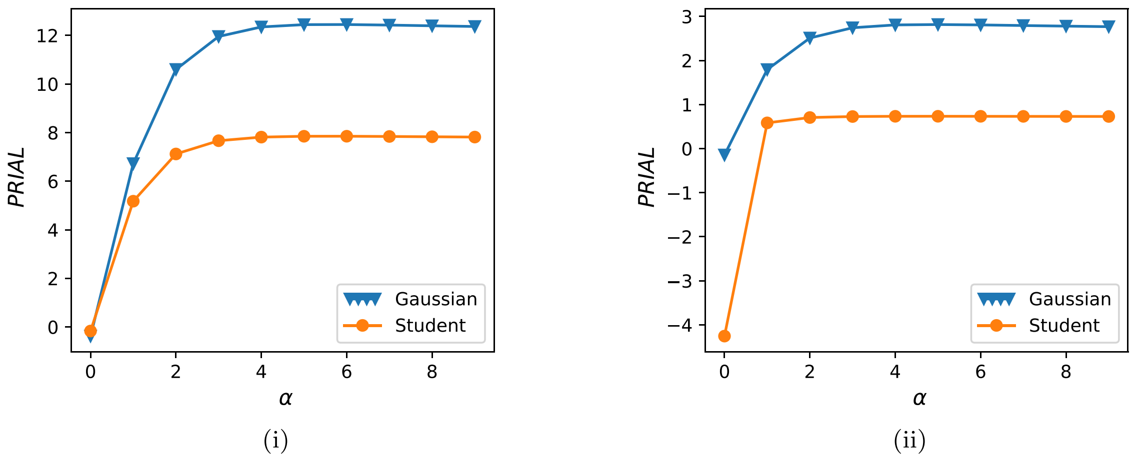

Fig. 2. PRIALS of $\hat{\Sigma}_{\alpha,b_0}$ under the data-based loss. The non-invertible case is considered, with $(p, m) = (50, 20)$, for the structures (i) and (ii) of Σ for the t-distribution, with $k = 5$, and the Gaussian distribution.

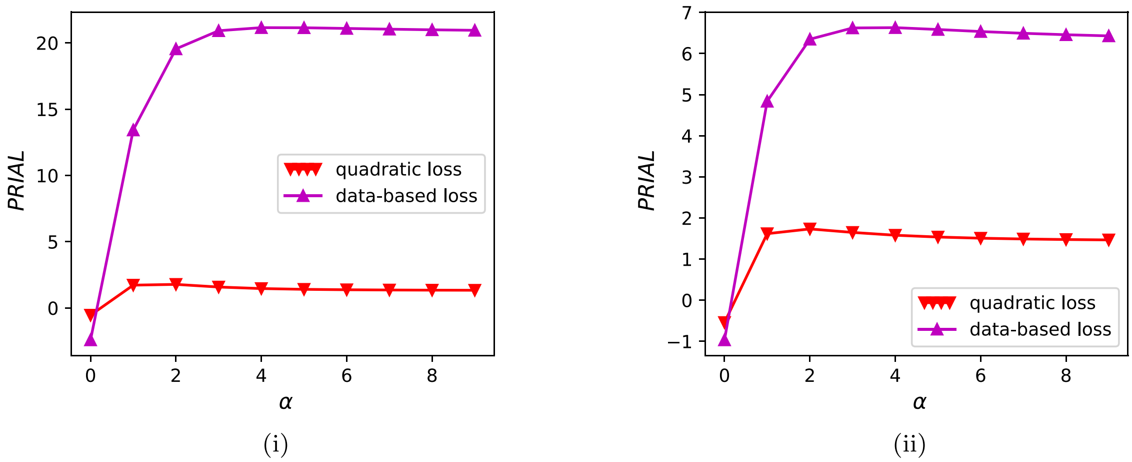

Fig. 3. PRIALS of $\hat{\Sigma}_{\alpha,b_0}$ under data-based loss and PRIALS of $\hat{\Sigma}_{\alpha,b_1}$ under quadratic loss. The non-invertible case is considered, with $(p, m) = (20, 10)$, for the structures (i) and (ii) of Σ under the Gaussian distribution.

4. Conclusion and perspective

For a wide class of elliptically symmetric distributions, we provide a large class of estimators of the scale matrix Σ of the elliptical multivariate linear model (1) which improve over the usual estimators $a\,S$. We highlight that the use of the data-based loss (4) is more attractive than the use of the classical quadratic loss (6). Indeed, (4) brings more improved estimators and their improvement is valid within a larger class of distributions. This means that (4) is more discriminant than (6) to exhibit improved estimators.

While, in our theory, our improved estimators depend on (Z, S), the dependence on Z is not exploited (this is in particular the case for our estimators in (19)). Recently, Tsukuma and Kubokawa (2016) considered, in the Gaussian case, alternative estimators depending on S and on the information contained in the sample mean Z, showing that an estimator for a mean could be used to improve the estimators of a covariance. In an elliptical setting, estimating a scale matrix by such estimators merits future investigations.

Acknowledgments

The authors would like to thank an anonymous referee for his/her valuable comments which helped to improve on a first version of the paper. All the authors are supported by the Franco-Algerian program PROFAS B+ 2016.

Supplementary data: Proof of Theorem 1 and its application conditions - Materials for the numerical study.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2021.109160.

References

Berger, J., 1975. Minimax estimation of location vectors for a wide class of densities. Ann. Statist. 3 (6), 1318–1328.

Candès, E., Sing-Long, C., Trzasko, J.D., 2013. Unbiased risk estimates for singular value thresholding and spectral estimators. IEEE Trans. Signal Process. 61, 4643–4657.

Canu, S., Fourdrinier, D., 2017. Unbiased risk estimates for matrix estimation in the elliptical case. J. Multivariate Anal. 158, 60–72.

Chételat, D., Wells, M.T., 2016. Improved second order estimation in the singular multivariate normal model. J. Multivariate Anal. 147, 1–19.

Díaz-Gacía, J.A., Gutiérrez-Jámez, R., 2011. On Wishart distribution: Some extensions. Linear Algebra Appl. 435 (6), 1296–1310.

Efron, B., Morris, C., 1976. Multivariate empirical Bayes and estimation of covariance matrices. Ann. Statist. 4 (1), 22–32.

Fang, K., Zhang, Y., 1990. Generalized Multivariate Analysis. 1990. Science Press, Springer-Verlag, Beijing.

Fourdrinier, D., Strawderman, W., 2015. Robust minimax Stein estimation under invariant data–based loss for spherically and elliptically symmetric distributions. Metrika 78 (4), 461–484.

Fourdrinier, D., Strawderman, W.E., Wells, M.T., 2018. Shrinkage Estimator. Springer.

Haddouche, M.A., 2019. Éstimation d'une matrice d'échelle (Thesis). Normandie Université; École Nationale Supérieure de Statistiques et d'Économie Appliquée (Alger), pp. 46–47, URL: https://tel.archives-ouvertes.fr/tel-02376077.

Haddouche, A.M., Fourdrinier, D., Mezoued, F., 2021. Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions. J. Multivariate Anal. 181, 104680.

Haff, L., 1980. Empirical Bayes estimation of the multivariate normal covariance matrix. Ann. Statist. 8 (3), 586–597.

James, W., Stein, C., 1961. Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, Berkeley, California, pp. 361–379.

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

Kubokawa, T., Srivastava, M.S., 1999. Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution. Ann. Statist. 27 (2), 600–609.

Kubokawa, T., Srivastava, M., 2001. Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. J. Multivariate Anal. 76 (1), 138–152.

Kubokawa, T., Srivastava, M., 2008. Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. J. Multivariate Anal. 99 (9), 1906–1928.

Srivastava, M.S., 2003. Singular Wishart and multivariate Beta distributions. Ann. Statist. 31 (5), 1537–1560.

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

Tsukuma, H., 2016. Estimation of a high-dimensional covariance matrix with the Stein loss. J. Multivariate Anal. 148 (C), 1–17.

Tsukuma, H., Kubokawa, T., 2015. A unified approach to estimating a normal mean matrix in high and low dimensions. J. Multivariate Anal. 139, 312–328.

Tsukuma, H., Kubokawa, T., 2016. Unified improvements in estimation of a normal covariance matrix in high and low dimensions. J. Multivariate Anal. 143, 233–248.

Tsukuma, H., Kubokawa, T., 2020. Shrinkage Estimation for Mean and Covariance Matrices. In: JSS Research Series in Statistics, Springer, Singapore.